21世纪 高等院校教材 生物科学系列

基础分子生物学

叶林柏 郜金荣 编著



科学出版社
www.sciencep.com





21世纪高等院校教材——生物科学类

基础分子生物学

叶林柏　郜金荣　编著

科学出版社
北京

内 容 简 介

本书以“内容新、重基础”为指导思想，从分子生物学的概念、术语、理论及新发现出发，系统介绍了分子生物学的理论。书中强调生物大分子的结构、构象与功能关系及DNA、RNA、蛋白质之间的相互作用，对DNA复制、基因表达调控、转录、翻译等内容给予详细介绍。

本书可作为生物学、医学、农学、药学等学科领域的本科生、研究生的教材，也可作为以上各领域的教师和研究工作者的参考书。

图书在版编目(CIP)数据

基础分子生物学/叶林柏，郜金荣编著. —北京：科学出版社，2004.8
(21世纪高等院校教材·生物科学类)
ISBN 7-03-012830-3

Ⅰ.基… Ⅱ.①叶…②郜… Ⅲ.分子生物学-高等院校-教材 Ⅳ.Q7

中国版本图书馆CIP数据核字(2004)第005877号

责任编辑：周 辉 彭克里 孙晓洁
责任校对：柏连海/责任印制：安春生/封面设计：陈 敬

科学出版社 出版
北京东黄城根北街16号
邮政编码：100717
http://www.sciencep.com
源海印刷有限责任公司印刷
科学出版社发行 各地新华书店经销
*
2004年8月第 一 版 开本：B5（720×1000）
2004年8月第一次印刷 印张：31
印数：1—3 000 字数：592 000

定价：38.00元

（如有印装质量问题，我社负责调换〈路通〉）

编写人员（以姓氏笔画为序）

孔令保　　叶林柏　　叶　力　　刘　静
吴剑媚　　郑　义　　郜金荣

插图制作人员（以姓氏笔画为序）

李宝宗　　吴正辉　　佘应龙
张　翼　　郜金荣　　廖庆姣



目录



第一章 绪 论

虽然 William Astbury 在 1945 年最先提出了“分子生物学”一词，但“分子生物学”这门学科的诞生时间是不能用具体年月来限定的。分子生物学本身就是一门技术和应用性较强的交叉学科。它有一个逐步积累的形成和发展的连续过程，它的逐步形成和发展，与生物学的其他学科以及化学和物理学的发展有着密切的关系。尤其是与生物化学、细胞学、遗传学和微生物学（包括病毒学）的发展关系更为密切。而分子生物学的发展又对其他学科的发展产生了重大影响，事实上，今天生物学的其他学科也发展到了分子水平，学科之间的互相交叉和渗透越来越广泛。分子生物学从人们决定打破细胞、研究组成细胞的大分子（如核酸、蛋白质、多糖）的结构与生物学功能开始，逐步在分子水平上认识了生物的遗传、变异、进化、遗传信息的传递、基因的表达与调控、细胞内和细胞间的信号调节、细胞分化和细胞癌变（增生）及凋亡、个体发育过程调控、感染、病理及机体防预，直到认识高等动植物乃至人类基因组学、蛋白质组学。这样，分子生物学的发展又回到了整体生物学，而生物学的各学科则可能在分子生物学的基础上统一为整体生物学或总体生物学。

一、分子生物学发展简述

早在 20 世纪 40 年代之前，人们就已经知道蛋白质和核酸是细胞内的重要成分，蛋白质参与或催化细胞内的化学反应，而且细胞组分可以在不依赖完整细胞的情况下进行反应从而实现物质转换，并且那时已经发现了病毒这种比细胞结构简单得多、微小得多的原始的生命形式，Stainly 可以把这种生命在试管中像化学物质一样“结晶”沉淀出来，这种生命只含蛋白质和核酸。但当时人们并不知道核酸是遗传物质，尽管摩尔根的“基因”概念这时已从“一个基因一种性状”向“一个基因一个酶”过渡，并且 Avery 的肺炎双球菌的转化结果已为人所知，但人们仍相信蛋白质是遗传物质，因为人们认为蛋白质是由 20 种不同的氨基酸通过肽键连接而成，比核酸只含 4 种碱基的结构要复杂得多。后来用从热灭活杀死的光滑型（野生型致病）肺炎球菌中提取的 DNA 转化活的粗糙型（不致病）肺炎球菌，证实可使部分后代恢复成野生型，仍有人怀疑这是由于提取的核酸中带有蛋白质所致。直至 1952 年 Hershey 在著名的搅拌试验中分别用 ^{32}P 和 ^{35}S 标记 T2 噬菌体的核酸和蛋白，证明在感染细菌时 T2 噬菌体只要把核酸注入细菌内，就能完成感染和复制的过程，从而产生子代噬菌体，这才最后确定了核酸是

遗传物质。现在我们都知道，DNA 和 RNA（RNA 病毒的基因组）都是遗传物质，是遗传信息的携带者。

确定 DNA 是遗传物质是分子生物学发展的重大里程碑，但是 DNA 只有 4 种不同的碱基，它如何能编码由 20 种不同氨基酸组成的蛋白质？它如何能准确复制从而把性状准确遗传给后代？它在核中携带的信息如何传递、如何准确控制细胞的各种生化反应？这些困惑科学家们的问题立即就被提出来了，要回答这些问题，必须首先了解 DNA 的结构。1953 年，Watson 和 Crick 借助于其他研究组提供的 DNA 晶体 X 射线衍射照片来研究 DNA 结构，提出了 DNA 双螺旋结构模型，这一模型令人相信 DNA 可以通过碱基配对的原则准确复制而把信息遗传给后代，也可以转录成副本从核中转运到细胞质。Crick 接着又提出“中心法则”，即遗传信息从 DNA→RNA→蛋白质。DNA 双螺旋结构模型的建立是分子生物学发展史上又一块丰碑，而有些学者认为这是分子生物学诞生的标志。此后，分子生物学发展越来越快，取得了不少重大突破性成果，取得的这些成果将永远载入学科发展史册，而这些成果同时也反映了分子生物学发展的历程。为使读者对分子生物学发展过程加深理解，现把这些重大发展略述如下：

1956 年　A. Kornberg 发现 *E. coli* DNA 聚合酶Ⅰ，1958 年分离该酶并在体外环境下酶促合成有活性的 DNA，因而于 1959 年获诺贝尔奖。

1958 年　Meselson 用著名的“密度转移”实验证实 DNA 的“半保留复制”；建立密度梯度离心技术。1968 年冈崎片段发现后提出 DNA 复制是半保留不连续复制。

1959 年　S. Weiss 发现转录酶。

1960 年　Jacob 和 Monod 经 10 余年研究后提出乳糖操纵子模型，这是第一个原核基因表达控制的模型，同时还预言 mRNA 的存在。

1961 年　S. Spiegelman 在 T2 感染的 *E. coli* 中发现 mRNA，建立分子杂交技术。

1965 年　R. W. Holley 测定酵母丙氨酸 tRNA 的一级结构，提出 tRNA 的“三叶草”结构模型。

1966 年　M. W. Nirenberg 和 H. G. Khorana 完成全部遗传密码的破译。

1967 年　Kates 和 McAuslan 发现真核转录酶，1970 年发现真核 mRNA 含有 polyA 尾巴（在天花病毒感染细胞中发现），于是用 oligo (dT) 柱分离纯化真核 mRNA 从而促进研究进展。

1968 年　M. Gellert 等五个实验室发现 DNA 连接酶，为发展体外 DNA 重组技术奠定了基础。

1970 年　H. M. Temin 和 D. Baltimore 同时发现不同反转录病毒的逆向

转录酶，补充了“中心法则”。

1970 年　H. O. Smith 发现了第一个 II 型 DNA 限制性内切核酸酶 *Hin*d II，导致之后一系列 DNA 限制性内切核酸酶的发现及应用，和 DNA 连接酶一起促进了 DNA 体外重组的发展。

1972 年　P. Berg 等三人建立 DNA 重组技术，建立了第一个体外 DNA 重组分子（λ dvgal DNA 片段克隆到 SV40）。并建立了含有哺乳动物激素基因的工程菌株，促进了 DNA 克隆技术的发展和应用。

1975 年　Furuichit 和 Miura 研究质型多角体病毒，发现真核 mRNA 中的 m^7Gppp 帽子结构。

1977 年　P. Sharp 和 P. Leder 分别从腺病毒 II 型和鸡卵白蛋白基因中发现真核基因内部含有内含子。

1977 年　Gilbert 和 F. Sanger 分别发明了不同的 DNA 测序技术，前者发明了化学断裂法；后者发明了加减法和聚合酶链式反应终止技术（即双脱氧终止技术），帮助人们研究基因的精细结构和排列乃至对人类基因组的研究。Sanger 因此第二次获诺贝尔奖（第一次因首次测定蛋白质——牛胰岛素的氨基酸顺序）。

1979 年　J. A. Shapiro 提出了描述转座子转座过程的 DNA 转座重组模型，即 Shapiro 模型。

到了 20 世纪 80 年代后，基因工程产业诞生，分子生物学发展到实用阶段，特别是 20 世纪 90 年代后，基因工程产业发展很快，已经可以用工程细胞株或菌株表达生产多种产品医药，包括疫苗、抗体、生长因子、细胞因子、激素、治疗用蛋白质或肽、基因治疗产品等。与此同时，转基因农作物、转基因动物技术也被用于动植物品种改良，使之具有抗虫、抗病、耐旱、耐盐、高产等更优良的品质，如抗虫转基因棉花、番茄、大豆等多种作物大面积的推广种植。克隆牛、羊的成功已成为家喻户晓的新闻。到上世纪末，产业的年利润增长速度已超过电子和信息产业，正在创造巨大的经济效益。生物技术行业将成为本世纪竞争最激烈的社会支柱产业。

除产业的形成和应用的发展外，20 世纪 80 年代后的基础研究发展，有人用“知识爆炸”来形容，这段时间的发展太快，重大成果太多而难以一一举出，例如 DNA 复制中的复制起始位点结构及参与复制的蛋白功能、互相作用；转录的起始及其调控；核酸酶发现及 RNA 的加工；mRNA 的翻译过程及调控；蛋白质的修饰、折叠和分子伴侣的发现；蛋白质结构与功能关系、蛋白质大分子相互作用；细胞表面受体及信号转导；细胞分化；癌基因与病变；DNA 定点突变技术、PCR 技术、单抗和人源抗体技术以及 30 亿 bp 的人类基因组全序列测定的完成等。这些研究成果，大大丰富了分子生物学的内容、加速了分子生物学的发展，

特别是人类基因组计划将会使人类更清楚地了解自身，对今后的医学及生命科学的发展必将产生极其深远的影响。今后分子生物学的研究方向，仍然会以蛋白质的功能研究为中心。蛋白质通过修饰、构象改变以及通过大分子间相互作用（蛋白质与核酸和蛋白与蛋白之间）形成大分子复合物而行使各种生物学功能如，基因表达的调控、发育和分化的控制等都是不同蛋白质的功能在时间和空间上的具体表现，而信号也是通过蛋白质修饰（如磷酸化或去磷酸化、与核苷酸或小分子如金属离子、NO 等的结合）、变构和相互作用来转导的。只有深刻了解蛋白质的功能，才能读懂基因组所包含的全部信息。

二、分子生物学的主要内容

分子生物学与生物学其他学科的交叉和渗透范围日益广泛，尤其是与生物化学、遗传学和细胞学之间的关系更加密切，以至于很难划分出明确的分界线，尽管它包括的范围很广，但它绝不能包括或代替其他生物学学科，就像细胞学不能包括遗传学和生物化学一样。有人认为生物化学已包括了分子生物学，把分子生物学看成是生物化学的一部分，这种观点也不对。生物化学的本质是“化学”，是生物体内通过酶催化、电子传递进行的化学反应（分解和合成反应）而促使机体内的新陈代谢、能量产生和消耗等，它的表示形式常常类似于化学反应式，涉及物质的转变，而通常发生质的变化，如淀粉被水解成葡萄糖，糖氧化反应等。而分子生物学涉及的是构象改变和大分子的互相作用，根本无法用化学反应式来表示，如调控蛋白识别并结合于某一基因的启动子序列而激活该基因表达，细胞表面受体的胞外部分与配体结合后引起胞内部分变构而进行信号传递等，所以分子生物学有它自已的内容和范围，它的主要内容如下：

1. 基因的精细结构及其表达调控

基因是指 DNA（或 RNA）上编码一种蛋白质或功能 RNA 的一段序列，基因的精细结构还应包括与基因转录和翻译调控有关的序列、转录起始有关区域的核苷酸序列。基因表达的调控决定蛋白质产生的时间和水平高低。基因表达是指基因所含的信息获得表现，因此一个基因的表达过程应包括：转录→RNA 加工→成熟 mRNA→转运→翻译→蛋白质加工和修饰→功能蛋白。这一过程的每一个步骤都受控制，例如一种转录通过 RNA 加工可以产生二、三种数量不相等的 mRNA，同时存在的数量相近的两种不同的 mRNA 翻译产生的蛋白质的量可以相差很大，即使在无细胞体系中，翻译也是如此，深刻研究了解这些过程中调控的细节是分子生物学研究的重要内容和艰巨任务。

2．信号传导

基因表达调控是在单个细胞内进行的，然而细胞如何与环境协调、细胞之间如何保持着一定的关系、多细胞多组织的复杂机体的生命活动如何保证高度协调、高度统一？这就涉及信号传导问题，包括神经信号和其他信号的传导。信号传导可以使机体对环境刺激迅速作出反应，信号命令某一群细胞定向分化并保持记忆而发育成某个组织或器官，也可命令细胞“凋亡”。如果信号传导失灵，机体的生命活动就会受到不同程度的影响，如发病甚至死亡。分子生物学对信号传导的研究、主要是研究受体结构、信号传导途径、各级联式反应有关的蛋白质及其作用以及信号导致的结果（如离子通道的开关或某一基因、蛋白的激活）。

3．蛋白质的结构、相互作用及其功能

基因的功能归根到底是通过蛋白质的功能而体现的，而蛋白质的功能又与它的构象有十分密切的关系，体内蛋白质之间（同种或不同种蛋白）又常常通过相互作用形成大分子体系（或蛋白质复合物）来行使功能。如何研究和确定蛋白质的结构是结构生物学的任务，而分子生物学研究则侧重于研究结构与功能之间的关系。体外研究蛋白质的结构，无论是用晶体的 X 射线衍射或核磁共振或冷冻电镜结合计算机分析的三维结构重建，获得的都是特定条件下的三维结构，而利用计算机的一些软件根据氨基酸序列预示的三维结构更是“仅供参考”。因为我们知道，蛋白质在不同条件下可以呈现不同的构象。例如 HIV 的两种糖蛋白 GP41 和 GP120，当 HIV 病毒与细胞受体结合时这些蛋白会发生结构的转变；又如疯牛病蛋白与正常蛋白比较，氨基酸序列并没有改变，但由于二级结构和构象的改变从而引起疯牛病。蛋白质的相互作用是广泛存在的，而在相互作用行使功能的过程中也常伴随着构象的改变，这一点我们在学习分子生物学时是值得注意的。

4．遗传、变异、进化的分子机制

DNA 或 RNA 如何复制、复制如何控制、DNA 损伤如何修复、DNA 变异和 DNA 重组的不同机制、生命的原始分子是什么及如何进化等也是分子生物学的研究内容。病毒的研究在 DNA 或 RNA 的复制及其调控方面做出了巨大的贡献，真核 DNA 的复制是比较复杂的，目前已知的参与复制的蛋白质至少有十几种，并且已经初步了解了这些蛋白在复制起始点形成复制叉的过程及在复制叉移动中的功能。但对于 RNA 基因组的复制认识还很缺乏。

5．分子生物学技术及其应用

用 DNA 重组技术使基因导入特定宿主中形成稳定的表达，以达到生产某种

产品、动植物品种改良和建立实验动物模型或基因治疗等目的也是分子生物学研究和学习的重要内容，而技术的发展往往也对基础研究起很大的推动作用，如前面提到DNA测序技术，分子克隆技术，近年来发展的mRNA差异显示技术、系列基因表达分析（SAGE）及芯片技术，如果读者对研究方法的原理和大致过程有了基本的了解，就能更容易理解一些理论问题。

三、如何学好分子生物学

分子生物学是一门技术性强、应用性强、发展迅速的课程。要掌握这门课程，需要有坚实的细胞学、生物化学、遗传学和化学基础，同时还需结合一定的实验课或有一定的研究经历。如果有了这些基础条件，本书还是可以自学掌握的。

第二章　生物大分子的基本结构和性质

一个典型的细胞中含有 $10^4 \sim 10^5$ 种不同种类的分子。经粗略地估计发现其中有一半是分子质量不超过几百道尔顿的无机离子和有机成分的小分子，而另一些分子则非常巨大，分子质量为 $10^4 \sim 10^{12}$Da，被称为生物大分子。这些大分子主要包括：蛋白质、核酸、多糖和脂类，它们分别是氨基酸、核苷酸、单糖和单脂的多聚体。由于这些大分子常常能以多种方式被修饰，因此，生物大分子还可以进一步分为亚类，如糖蛋白（蛋白质携带有糖基）、脂蛋白（蛋白质携带有脂类）、脂多糖（多糖携带有脂）和糖基化核酸（核酸的碱基上携带有糖）。

生物大分子执行各种功能。例如，核酸储存和携带遗传信息；多糖提供能量并且组成植物和许多微生物的细胞壁；脂蛋白是组成细胞膜的主要成分，它是细胞维持内部环境的重要成分，并负责运输物质进出细胞；最通用的生物大分子——蛋白质，可以催化化学反应、调节信息的流量、也是许多结构亚基的成分。

有关生物大分子性质方面的知识，对了解生命活动的过程是必须的。

第一节　生物大分子概述

一、生物大分子的化学结构

1. 蛋白质

蛋白质是由若干个聚氨基酸（多肽）组成的多聚体，一个氨基酸可以看做是与一个羧基、一个氨基和一个侧链（R）连在一起的单个碳原子（α 碳原子），这些侧链一般是碳链或碳环。多种功能基团可与这些侧链相接触。最简单的侧链是甘氨酸的羟基和丙氨酸的甲基。组成蛋白质的 20 种氨基酸中，只有脯氨酸在基本化学结构上不同于一般的氨基酸。在脯氨酸中，氮原子包括在环中，恰当地说，脯氨酸是一种亚氨基酸。由于脯氨酸的特殊结构，使它可以引导多肽链拐弯，从而使接近脯氨酸的地方明显地影响蛋白质的三维结构。

一个氨基酸的氨基可以与另一氨基酸的羧基反应形成二肽，两个氨基酸之间形成的这种键叫肽键。多个氨基酸被连续地连接在一起，形成一条线状多肽链。一般肽键数目超过 15 个时，此多肽就可以称为蛋白质。蛋白质也可由几个多肽链聚集而成。一个蛋白质是一个多聚体，在该多聚体中，碳原子和肽单位交替形成一个带侧链的有序排列的线状链，这个线状链称为该分子的骨架。每个蛋白质

分子的两个末端的性质是截然不同的，一端有一自由的—NH_2，称为N端，另一端有一自由的羧基，称为C端，这两个不同的末端也常被分别称为N末端和C末端。

氨基酸侧链之间常以非共价键结合，但半胱氨酸的—SH基是个例外。—SH常与另一个相同或不同的多肽链中半胱氨酸的—SH反应形成二硫键（S—S），这对决定一个蛋白质的三维结构很重要。许多蛋白质含有其氨基酸侧链基团复合物中螯合的金属离子，如：Zn^{2+}、Cu^{2+}和Fe^{3+}，这些金属离子常常结合在组氨酸和谷氨酸上。

2. 核酸

核酸是核苷酸的多聚体。每个核苷酸均含有以下三种成分。

环状五碳糖，有核糖和脱氧核糖，核糖存在于核糖核酸（RNA）中，脱氧核糖存在于脱氧核糖核酸（DNA）中。这两种核糖的区别在于脱氧核糖2′-碳原子上的—OH基被脱了氧，只剩H。这个区别使得DNA比RNA更具化学稳定性。

核糖的1′-碳原子上通过N-糖苷键连接有嘌呤或嘧啶碱基。RNA中有四种碱基，A（adenine）、U（uracil）、G（guanine）和C（cytosine），DNA中的四种碱基是：A、G、C和T（thymine）。RNA和DNA都含有A、G和C，U只在RNA中，而T只在DNA中，但此规则也有极罕见的例外，如一些tRNA分子中含有T，而极少数噬菌体，其DNA中含U而不含T。

核糖的5′-碳原子上通过磷酸二酯键连接有磷酸基团，这些基团使核酸和核苷酸携带强大的负电荷。

核糖连接上碱基就叫核苷（nucleoside），核苷再磷酸化后叫核苷酸（nucleotide）。核酸中的核苷酸的连接方式是一个核苷酸的5′-磷酸和另一核苷酸中的3′-OH形成第二个磷酸酯键而共价连接，3′和5′-碳原子上的磷酸都是酯化的，这样的单位常被称为磷酸二酯基团（phosphodiester group）。

嘌呤和嘧啶碱基彼此之间不形成任何共价键，因此，一个多核苷酸含有一条糖与磷酸交替出现的骨架，这一骨架具有一个3′-OH末端和一个5′-P末端。在实验室里可制备3′-P末端和5′-OH末端的多核苷酸。

3. 多糖

多糖是糖（最常见的是葡萄糖）或糖的衍生物的多聚体，多糖是很复杂的分子，因为有时在许多对碳原子之间可能会产生共价键，这样就产生了一个糖单位与两个以上的其他糖单位连接在一起的效应，结果将会形成高度分支的大分子，这些分支结构有时很庞大，甚至用显微镜就可以观察到。例如，许多细菌和植物细胞的细胞壁就是由单个庞大的多糖分子构成的。

4. 脂类

许多脂类由较小的单位组成，它们参与许多大分子的装配。自然界中的脂类是难溶于水的碳氢化合物，它们可以储存或转移能量，并且是组成膜、保护性外鞘及其他细胞结构的重要组分。甘油酯是由一个到三个饱和或不饱和脂肪酸与1分子甘油以酯键相连而成。动物中的脂肪中包含的饱和脂肪酸是固体形式，而植物中的是不饱和脂肪酸，是液体形式。膜中含有磷脂、鞘脂及鞘磷脂。

二、决定蛋白质和核酸三维结构的非共价相互作用

大分子的生物学性质主要是由非共价的相互作用来决定的。这些作用的结果使每个分子获得独特的三维结构。

1. 无规则线团

线状多肽链和核酸链含有多种键，这些键中，有的可以自由旋转，在不存在任何链内相互作用的情况下，每个氨基酸或核苷酸将是自由旋转的。这种自由旋转只受到原子不能占据同一空间的限制。像这样一种链的三维构象就叫无规则线团（random coil），它们是稍有点紧密的球形结构，这种结构由于受到溶剂分子的不停的碰撞而不断改变其形状。

核酸和蛋白质在天然状态下几乎不存在无规则线团，因为组成核酸和蛋白质中的这些链的各因素之间存在许多相互作用。这些相互作用包括氢键、疏水相互作用、离子键以及范德华作用等，例如，多肽链具有所有这四种类型的相互作用，又如，核酸的碱基之间可通过氢键和疏水作用相互吸引，但如果它们具有同一种电荷，它们之间就会互相排斥。

2. 氢键

生物系统中常见的氢键如下：

① C=O···H—N　② —C—OH···O=C　③ N—H···N

其中的氢键是由两条链之间两个配对的核苷酸碱基之间形成的。在双链DNA中，氢键相互作用是维持DNA双螺旋结构的主要因素。在蛋白质中连接在一个肽键上的一个亚氨基上的氢原子和相邻的连接在不同肽键上的氧原子之间的氢键相互作用可以产生好几种特定的多肽构象。

3. 疏水相互作用

疏水相互作用是指两个微溶于水的分子或分子的某些部分之间的相互作用。难溶于水的两个分子（可能是两个不同的分子）趋向于缔合。这个现象可以简单地解释为，水分子相互形成氢键，一个分子如果能与水形成氢键，该分子就可溶于水。而相互形成氢键的水分子可以在不溶于水的分子周围形成一层“壳”(shell)，水的有序化在热力学上降低了熵值。两个不溶于水的分子将分别形成两个这样的“壳”，如果这两个分子接触，那么单个水“壳”将围住这两个分子，由于严格的几何学原因，这个围绕一对分子的水“壳”将是围绕一个分子的那两个水“壳”所需有序水分子的数目的将近二倍。一般来说，如果微溶于水的分子相接触，那么，每个溶质分子的有序水分子的平均数目总是比原来少的。因此，微溶于水的分子成簇是符合热力学原则的。这种成簇的趋势叫做疏水相互作用。在疏水作用中是没有键形成的，只是更有可能成簇，或者换句话说，所有的分子包括溶剂和溶质的无序排列的熵值是比有序排列的熵值大的。

核酸和蛋白质的许多成分都有疏水性质，例如，核酸的碱基是局部携带弱电荷的平面有机环，这些局部的弱电荷足以维持溶解性，可是，那些大的难溶于水的有机环部分会引起水分子有序化，因此碱基趋向于成簇，最有效的成簇类型是碱基堆积。因碱基在链中是相邻的，所以堆积使多聚核苷酸单链刚性，使之趋向于伸展态而不至于形成无规则线团。碱基堆积是决定核酸结构的重要因素。许多氨基酸侧链是难溶的，这会引起苯丙氨酸的苯环堆积，丙氨酸、亮氨酸、异亮氨酸和缬氨酸中烃链将形成紧密的非堆积簇：由于这些氨基酸不一定是相邻的，因此疏水相互作用会趋于把一个多肽链上远离的疏水部分聚到一起。

4. 离子键

离子键是不相同的电荷之间相互吸引的结果，若干氨基酸侧链已经离子化，带负电荷的羧基基团（门冬氨酸和谷氨酸）和带正电荷的氨基基团（赖氨酸、门冬酰氨和精氨酸）可以形成离子键，也趋向于将链中远离部分聚到一起。离子相互作用也可以是排斥的，如两种相同的电荷之间。离子键是非共价相互作用中最强的一种，然而离子键可被极端 pH 所破坏，因为 pH 可改变基团的电荷，同时，离子键也可以被高浓度盐所破坏，这些盐离子将带电基团保护起来，使它们之间不能形成离子键。

5. 范德华引力

范德华引力存在于所有分子之间，它们是长期偶极子和电子环流的结果。两个原子之间的吸引力相当于 $1/r^6$，其中，r 是两原子核之间的距离，所以这种吸引是很弱的力，而且只有当两个原子核很近时（1～2Å）才会明显，如果两个

外电子层重叠，那么将会有一个强有力的推斥力起作用。范德华半径（Van der-Waals radius）是吸引力和排斥力精确地达到平衡时的两原子核之间的距离。原子的范德华半径各不相同，如：

原子	半径（Å）
H	1.2
O	1.4
N	1.5
S	1.85
P	1.9
C	2.0

两个原子之间的相互作用常常不足以抵抗原子的热运动能，但如果一个分子中两个分开的区域形状准确互补，那么，多个原子的若干部分之间的范德华相互作用就可以将它们保持在一起。维持一个分子的三维结构常常是多种弱的相互作用同时起作用。

三、研究生物大分子的基本方法

1. 速度沉降

沉降系数（sedimentation coefficient，S），大多数大分子的 *S* 值为1×10^{-13}～100×10^{-13}s之间，因为 Svedberg 发明了超离心，所以将 10^{-13}s 叫做 1 个 Svedberg，即一个 S。这在分子生物学中被普遍应用，例如，一个分子的 *S* 值是 30 个 Svedberg，该分子就是一个 30S 的分子。

2. 区带离心

区带离心（zonal centrifugation）是用蔗糖，偶尔也用甘油或其他溶液制备梯度，也常被称为蔗糖梯度离心（sucrose graduation centrifugation）。

3. 平衡离心

在密度梯度中平衡离心（equilibrium centrifugation），用 CsCl 制成梯度，离心至平衡时，不同密度的分子就停留在自己的密度处，形成不同的带，因而可以分离不同的分子。

4. 电泳

分子生物学中用的最多的方法是凝胶电泳（elecectrophoresis）。

5. 电镜观察

用电镜（electronmicroscopy）可以观察某些生物大分子并测定这些生物大分子的相对分子质量。

四、生物大分子的分子质量测定

1. DNA 的分子质量测定

凝胶电泳法最适合用于分子质量小于 5×10^6Da 的 DNA 分子质量的测定。该方法是用已知分子质量的 DNA 分子作为标准样品，电泳后根据分子质量和迁移率的关系，做出标准曲线，再根据目的样品 DNA 的迁移率，算出其分子质量。

中等大小的 DNA 分子的分子质量（$5\times10^6\sim100\times10^6$Da）测定最好用电镜法。根据双链 DNA 分子的长度和分子质量的关系（1μm 相当于 2×10^6Da），可在电镜照片上量出分子长度再计算其分子质量。

大的 DNA 分子的分子质量（$>100\times10^6$Da）测定最好用速度沉降法，即测其 S 值。

2. 蛋白质的分子质量测定

对蛋白质的分子质量测定最方便的方法是 SDS-聚丙烯酰胺凝胶电泳（SDS-PAGE）。

第二节　DNA 的结构和性质

脱氧核糖核酸（DNA）是活细胞中最重要的分子，它含有特定细胞的全部遗传信息。而核糖核酸（RNA）除了作为某些病毒的遗传物质外，主要是执行其他一些重要的生物学功能。

一、DNA 的基本结构

1. 双螺旋结构是 DNA 的基本结构

早期对 DNA 的研究和各种实验表明，DNA 分子是一个长的高度有序的结构。通过 X 射线衍射分析，得知了关于 DNA 分子的各个部分的大小和排列，观察到分子是螺旋形的，核苷酸碱基是平面堆积，两个碱基之间的距离是 3.4Å。化学分析证明，DNA 分子含腺嘌呤（adenine，A）、鸟嘌呤（guanine，C）、胸腺嘧啶（thymine，T）和胞嘧啶（cytosine，C）。四种碱基的比是 A = T，G = C。James Watson 和 Frances Crick 结合物理和化学数据，根据 X 射线衍射照片，设

想 DNA 分子是两条链互相缠绕在一起，形成一个双螺旋。在这个模型里，核糖-磷酸的骨架在分子的外侧按螺旋方式排列，碱基则伸向双螺旋的内部。一条链的碱基与另一条链同一平面上的碱基通过氢键形成嘌呤-嘧啶碱基对，即 A-T 和 G-C。由于碱基对中均含有双环的嘌呤（A 和 G）和单环的嘧啶（T 和 C），所以，每个碱基对的长度是接近一致的。每个碱基对中的碱基在同一平面上，并且与双螺旋的轴垂直。相邻的碱基对以 36°旋转，每个螺圈有 10 个碱基对，双螺旋两条链之间的距离是 20Å，分子质量为 2×10^6Da/μm。

DNA 双螺旋有两个槽沟，一个大沟（major groove）和一个小沟（minor groove），这些槽沟足以允许蛋白质分子进入，与碱基接触。碱基配对是 DNA 的一个重要性质，它意味着两条链的碱基顺序互补，这就暗示了 DNA 的复制机制，因为复制时只要合成互补链就行了。

DNA 双螺旋的两条链是反向平行的（图 2-1）。反向平行有两重含义，其一是双螺旋分子的每一端都有一 3′-OH 末端和一 5′-P 末端，其二是两条链的走向不同，双链的走向不同是指两条链上的糖环的方向不同。一般书写时，5′-P 在左边，如 Patcg-3′-OH，也常写为 pApTpCpG……。

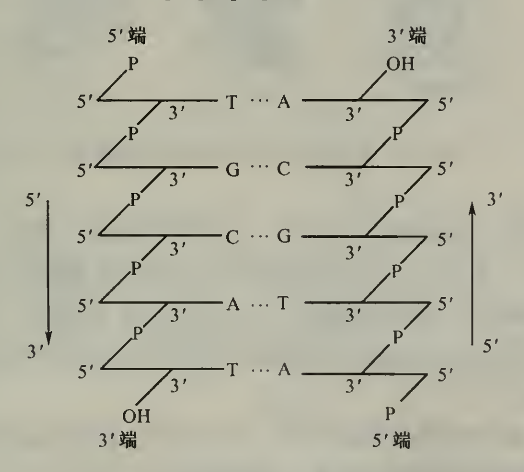


图 2-1　DNA 的两条互补链反向平行

不同生物 DNA 碱基组成不同，测定某种生物的碱基组成通常所用的方法是测 G+C/总碱基的碱基比率。这样测定的 G+C 含量对高等生物来说一般变化不大，在 50%左右，不同种之间变化在 49%～51%。而对低等生物，G+C 含量变化很大，如细菌中，滕毛属梭状芽孢杆菌（*genus clostridium*）G+C 含量极低，只有 27%，而滕毛属八叠球菌（*genus sarcin*）G+C 含量达 76%。大肠杆菌中 G+C 含量为 50%。用 CsCl 平衡梯度离心可以很便利地测定一种生物 DNA 的 G+C 含量，因为在 CsCl 溶液中，DNA 的密度与 G+C 的含量成线性关系。

一般天然的 DNA 都是经典的 Watson-Crick 结构，即右手螺旋结构，但在某

些条件下，DNA 分子具有左手螺旋形式，在天然条件下 DNA 分子中也有某些区域明显是左手螺旋。

湿度不同对 DNA 的结构也有影响。在相对湿度为 66%时，DNA 为 A 型。相对湿度为 92%时，DNA 为 B 型。A 型 DNA 每螺圈为 11 个碱基对，碱基平面不是垂直于双螺旋的轴而是倾斜 20°。B 型 DNA 是右手螺旋即经典的 Watson-Crick 结构。左手螺旋 DNA 即 Z-DNA 每螺圈有 12 个碱基对，而且 Z-DNA 双螺旋只有一个槽沟（图 2-2)。

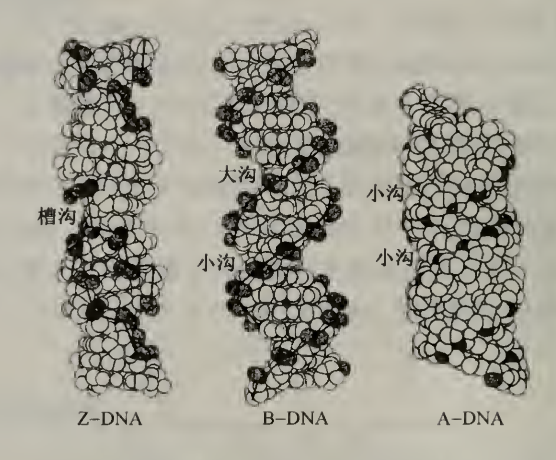


图 2-2　核酸可以形成几种形式的螺旋

不同种类的生物，基因组大小不同。在病毒和原核生物中，某种病毒或细胞的整个基因组是由一条 DNA 分子缠绕而成（RNA 病毒除外)。在真核生物中，包括单细胞生物如藻类、酵母和原虫，它们的 DNA 是组成一定数目的染色体，每条染色体由一条巨大的 DNA 分子组成，不同 DNA 分子的长度范围从 $0.7\mu m$ 到 $40\ 000\mu m$（4cm)。

一般来说越复杂的生物基因组越大，越简单的生物其基因组越小（但 toad 蟾蜍和南美洲肺鱼基因组比人的还大，这是例外)。

DNA 分子的长和宽的比是 $10^7:1$，所以很细很长，因此在分离提取 DNA 时，移液、倾倒溶液和混合溶液均易使 DNA 分子断裂，操作要特别小心。

从细菌中提取的 DNA 总是断裂的，断裂后的平均分子质量为 2.5×10^6Da，从高等动植物中分离的 DNA 不断裂的现象很少见。用一标准 DNA 和提取的 DNA 同时电泳，可以检查提取过程中 DNA 断裂的程度。

2. 决定 DNA 结构的因素

在同一条链中，DNA 的螺旋结构由碱基堆积决定，DNA 的双螺旋结构由配对碱基间的氢键决定。以下实验结果可以证明这些结论。

(1) 变性和熔解曲线

核酸对紫外光的吸收量由其结构决定，分子排列越趋于有序，吸收紫外光的量越少，因此可以利用紫外吸收的性质来测定 DNA 的含量。单链、双链和自由碱基的 A_{260}吸收值不同。50μg/ml DNA 溶液测得的 A_{260}吸收值分别为：双链 DNA $A_{260}=1.00$；单链 DNA $A_{260}=1.37$；自由碱基 $A_{260}=1.60$。

双链 DNA 缓慢加热，其溶液对紫外光的吸收值增加，叫增色效应（hyperchromic），而单链 DNA 缓慢降温，其对紫外光吸收值减少，叫减色效应（hypochromic）。加温使 DNA 变性时，其紫外光吸收值达最大值的一半时的温度叫熔链温度（melting temperature，T_m）。

将双链 DNA 缓慢加热，不同时间取样测 A_{260}吸收值可制得该 DNA 的熔解曲线（图 2-3）。这种曲线反映了随着 DNA 结构的变化，其对紫外光的吸收会不断增加。说明加热破坏了原来维持 DNA 双螺旋结构的因素（氢键和碱基堆积）。

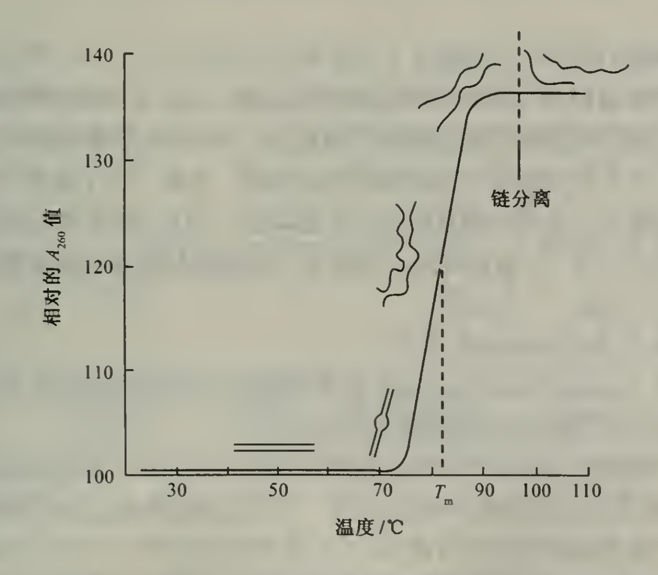


图 2-3　DNA 分子的 T_m 值和不同程度解链的 DNA 分子的可能构型

(2) DNA 中存在氢键的证据

从不同生物中分离的 DNA 其 G+C 含量差异很大，可以从 20%到 80%。用这些不同的 DNA 做变性曲线，发现 G+C 含量高的，其 T_m 值也高。已知 DNA 双螺旋的稳定因素是碱基对中的氢键，由于 G、C 之间是三个氢键，而 A、T 之间是两个氢键，所以打破 G、C 之间的氢键比打破 A、T 之间的氢键需要的能量要高，因此 G、C 含量高的 DNA 其 T_m 值也高（图 2-4）。

尿素（ $H_2N-\overset{\overset{O}{\|}}{C}-NH_2$ ）和甲酰胺（ $H_2N-\overset{\overset{O}{\|}}{C}-H$ ）这两种物质加到 DNA

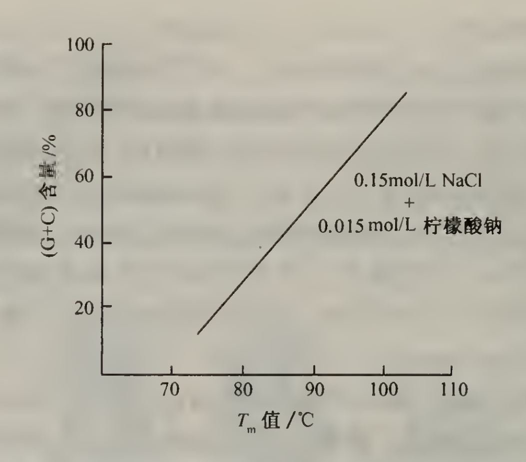


图 2-4　不同 DNA 分子的 G+C 含量及其对应的 T_m 值

溶液中可以降低 DNA 的 T_m 值。

这两种物质可以和 DNA 中的碱基形成氢键。实际上 DNA 中的碱基对之间的氢键是处于断开和重新形成的动态平衡状态。在 DNA 溶液中加入上述试剂的一种，即可与处于氢键断开状态的碱基形成氢键，因此，这个碱基不能再与另一条链中的碱基配对，这部分碱基即处于熔链状态，因而 DNA 变性处理时所需要输入的能量就少，其 T_m 值就降低。事实上，所有能和碱基形成氢键的物质均能降低 DNA 的 T_m 值。

（3）DNA 中存在疏水相互作用

疏水反应（hydrophobic reaction）是两种微溶于水的物质或疏水物质分子间的反应。DNA 分子间的反应就是疏水反应。

如果一种溶剂能加强弱的可溶性物质与水的相互作用，或能打破该物质外面的“水壳”，那么就会降低疏水相互作用，则 T_m 值也降低。如甲醇能增加碱基的溶解度，三氟醋酸钠盐可以打破 DNA 分子外的“水壳”，当在 DNA 溶液中加入上述任何一种物质时，都能使 DNA 的 T_m 值大大降低（图 2-5）。因此推断疏水相互作用也是关于 DNA 结构稳定的一个重要因素。

（4）碱基堆积

用旋光色散（opticalrotatorydispersion，ORD）和圆二色（circulardichroism）技术分析单链 DNA 和 RNA，证明其碱基排列在螺旋里，即使在一个二核苷酸里，碱基也是平行排列的，所以推断即使在一个小小的 DNA 分子中，也存在碱基堆积，双链 DNA 中的碱基堆积比单链 DNA 中要多。由于双链 DNA 中有氢键，氢键和碱基堆积是相辅相成的，从而使 DNA 结构稳定。

碱基堆积本身也是一种协同作用。在碱基堆积中，处于中间的碱基比处于两边的碱基要稳定，两边的碱基越多，中间的碱基越稳定。线状双链 DNA 分子的

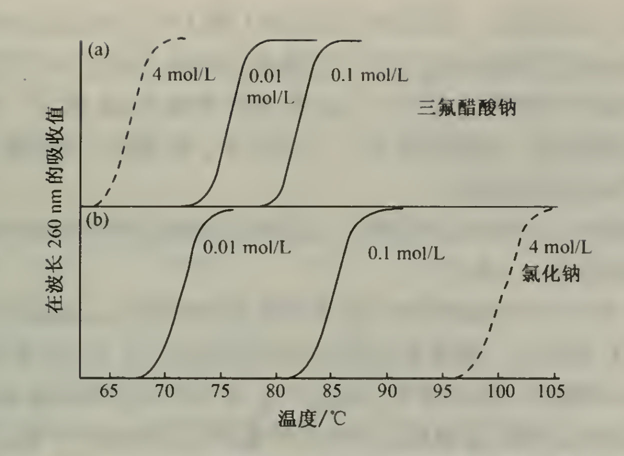


图 2-5　氯化钠和三氟醋酸钠对 DNA 溶链温度的影响对比

两端常没有氢键，有 7 个没有配对的碱基处于游离状态。少于 15bp 的双链 DNA 片段，其 T_m 值特别低。只含 3bp 的 DNA 片段在室温下不稳定。中间只有很短的配对区而两侧是非配对区的 DNA 分子在生理条件下不能保持其构型（图2-6）。

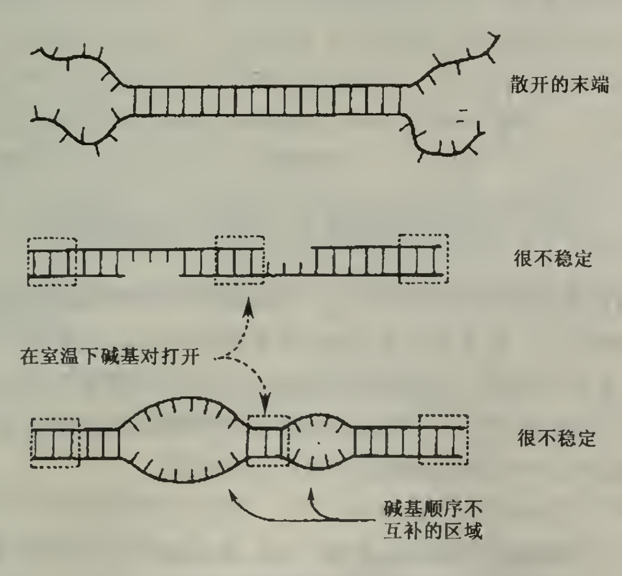


图 2-6　碱基堆积的协同作用

（5）溶液的离子强度对 DNA 结构的影响

DNA 分子两条链之间及其相邻的碱基之间有吸引力，此外，由于 DNA 分子两条链上的 PO_4^{3-} 带负电荷，所以 DNA 分子之间存在排斥力，这些负电荷如不被中和，强大的排斥力将使链分离。例如，DNA 的 T_m 值由于缓冲液的离子浓

度不同而不同。在室温下，随盐离子浓度的下降 DNA 的稳定性会降低，其 T_m 值也降低。DNA 在双蒸水中室温下可以变性。加入 NaCl，Na^+ 可以中和 PO_4^{3-} 的负电荷，使其不再排斥，因而一定的离子强度使 DNA 稳定。在生理盐水中 DNA 就处于这种状态。在高盐浓度下，DNA T_m 值升高。高盐降低了碱基的溶解度，增加了疏水相互作用。

很多证据证明，DNA 结合阳离子，如 DNA 的钠盐与 DNA 的铯盐分子质量不同：NaDNA/CsDNA＝0.75。

生理条件下的 DNA 双链常处于打开而形成单链的 breathing 状态。甲醛能和碱基上的—NH_2 起反应，使碱基之间不再形成氢键。在 DNA 溶液中加入甲醛，可以引起 DNA 缓慢的不可逆变性（图 2-7）。与甲醛反应的氨基必须处于自由状态，这说明了碱基之间的氢键是处于断开和重新形成的动态平衡之中。breathing 结构多出现在 A·T 富集区而不出现在 G·C 富集区。

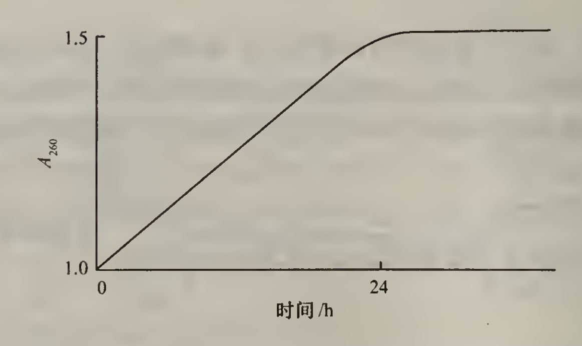


图 2-7　在 4%的甲醛溶液中，DNA 的 A_{260} 光吸收增加

DNA 溶液的某些物理性质的改变，例如黏度和旋转偏振光能力的降低等，随着变性过程的发生，常常会发生 DNA 的氢键和疏水作用消失、DNA 螺旋结构受到破坏且分子失去刚性。这些规则结构的瓦解常会伴随着双链的完全解开。过去人们曾认为 DNA 分子太长，完全把螺旋松开使两条链分离是不可能的，但很多证据清楚地表明 DNA 的两条链确实可以完全分离。

制备一条 ^{15}N ^{14}N 杂交双链 DNA 分子，该分子的密度等于 ^{14}N ^{14}N DNA 和 $^{15}N^{15}N$ DNA 分子密度的平均值，把该 DNA 溶液加热到不同温度，测 A_{260} 值，得到熔链曲线，将每个样品进行 CsCl 密度梯度离心，发现 A_{260} 值恰好高于最高值的样品在 CsCl 溶液中其 DNA 两条链是分开的有两条密度带（图 2-8）。在 37℃条件下，链分离的速度为 12×10^7 碱基/s。细菌 DNA 复制速度是 10^6 碱基/s，所以链分离不会限制 DNA 的复制速度。

在某些条件下，DNA 可以变性。有许多蛋白质能使 DNA 解旋，这些蛋白质被称为松旋蛋白（relaxation protein）、螺旋降稳蛋白（helix-destabilizing protein）

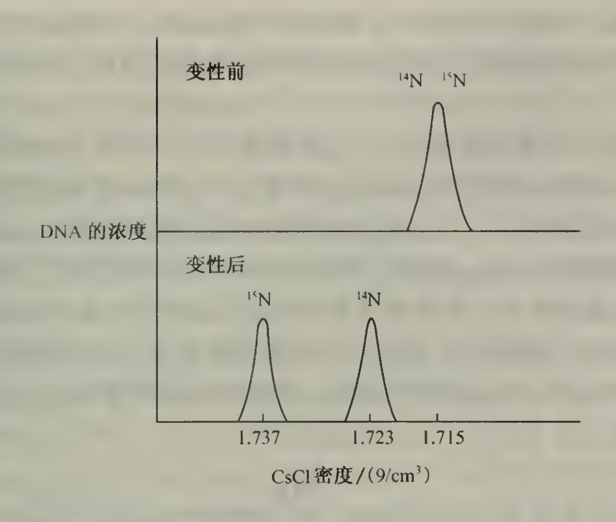


图 2-8　通过 CsCl 密度梯度离心，^{14}N ^{15}N DNA 变性后 ^{14}N ^{15}N 链分离成两条带，变性前为一条带

或熔链蛋白（melting protein）。其中了解得最清楚的是一种最简单的蛋白——螺旋降稳蛋白。由于最早发现是由大肠杆菌 T4 噬菌体 *32* 基因编码的，所以称 32 蛋白。32 蛋白有两个重要特点可以使 DNA 变性，第一，它能与 DNA 单链紧密结合；第二，蛋白质分子之间有协同效应。由于这两个特点，32 蛋白能线性地排列在 DNA 单链上并与之紧密结合（图 2-9）。一旦有一个 32 蛋白结合到 DNA 的 breathing 处，它就能与几个碱基结合。这不仅给其互补链提供了结合 32 蛋白的机会，而且使它自身结合链上相邻的几个碱基失去稳定性，32 蛋白不断加入，直到 DNA 分子全部变性。自身 32 蛋白对于 DNA 复制时的解链和遗传重组中的单链入侵都是必需的。

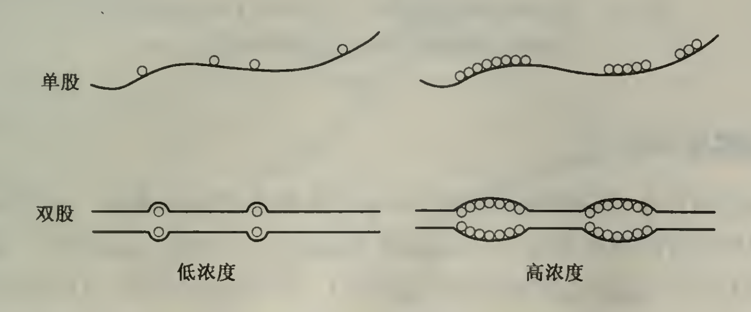


图 2-9　32 蛋白与 DNA 结合后使 DNA 变性

DNA 在碱性条件下可以全变性。DNA 的螺旋结构靠碱基堆积维持，氢键则把两条链拉在一起。加热可破坏上述结构，从而使 DNA 变性。但高温可能也会

破坏磷酸二酯键，故加热可能会产生断裂的单链混合物。在高 pH 时，改变配对碱基的某些基团的带电情况，可以使氢键的形成受到阻碍，从而使 DNA 全变性。

DNA 在 90℃以上通常全变性，A_{260}值增加 37%。变性 DNA 溶液中的 DNA 全部是单链，这些单链堆积力降低。当变性的 DNA 溶液降温到室温，在 0.05mol/L 左右的盐浓度下，其 A_{260}值比变性温度时达到的 A_{260}值明显下降，一般会降低到天然 DNA A_{260}值的 1.12 倍。冷却后大约有三分之二的碱基重新形成氢键或在链内碱基堆积从而形成紧密结构。在低盐（<0.01mol/L）浓度下，分子呈展态，很少有氢键形成（图 2-10）。如果在变性 DNA 溶液中加甲醛或保持其高 pH，则可保持 DNA 的变性状态，即使在高盐和室温下亦无链内氢键的形成和碱基堆积。

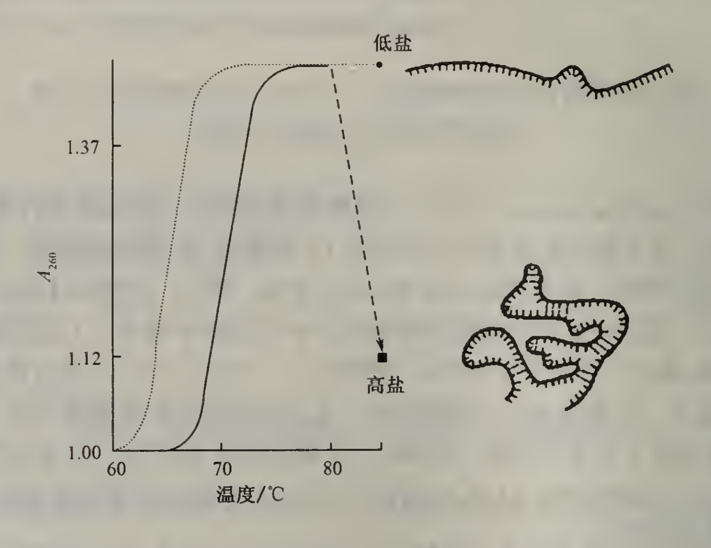


图 2-10　链分离后降温到 25℃，DNA 分子在高盐溶液中链内形成氢键，在低盐溶液中呈展态

3. 环状超螺旋 DNA

大多数原核生物及病毒的完整 DNA 分子都是双链环状的。一个环状 DNA 分子可以是共价闭合环 DNA（covalently closed circle DNA，cccDNA），也可以是切口环状 DNA（nicked circle DNA，ncDNA），在 DNA 分子中，打开一个磷酸二酯键就产生一个切口（nick）。共价闭合环状 DNA 分子常常形成超螺旋（superhelix）或者称其为超级线团（super coil）；超螺旋的结构有两种，一种是负超螺旋，另一种是正超螺旋。

负超螺旋（negative superhelix）：形成超螺旋时，旋转方向与 DNA 双螺旋方

向相反，旋转的结果使 DNA 分子内部张力减少，称其为松旋效应（under-wound）。在自然条件下的共价闭合环状 DNA 呈负超螺旋结构。

正超螺旋 DNA 与负超螺旋相反，形成超螺旋时的旋转方向与 DNA 双螺旋方向相同，结果加大了 DNA 分子内部张力，称其为紧旋效应（overwound）。

超螺旋结构中存在单链区，这些单链区 90% 以上为 A-T 区，它与起始遗传重组、起始 DNA 复制和起始 mRNA 的合成有关。

用在碱性 pH 下沉降的方法和在含溴化乙锭的 CsCl 中平衡离心的方法可以对共价闭合环状 DNA 进行实验测定。

在 pH 大于 11.3 时；所有氢键均断裂，DNA 分子松旋，如果是一个切口环状 DNA，那么变性后将产生一个线状分子和一个单链环分子，该线状分子的 S 值比此单链环状分子的大 14%。而单链线状 DNA 分子的 S 值比其天然状态的大 30%（在 0.3mol/L 或大于 0.3mol/L 的盐溶液中）。共价闭合环状 DNA 分子由于其两条链不分开，因此其 S 值比天然线状 DNA 分子的 S 值大两倍。

溴化乙锭（ethidium bromide）能与 DNA 紧密结合，并使 DNA 的密度降低至接近 $0.15g/cm^3$，溴化乙锭插入到 DNA 分子碱基对之间，当插入较多时，会引起 DNA 分子松旋；切口环状和线状 DNA 分子由于不存在松旋的拓扑学问题，就能够比共价闭合环状 DNA 分子结合更多的溴化乙锭分子。用足够量的溴化乙锭处理这三类 DNA 分子后进行密度梯度离心，因为共价闭合环状 DNA 分子比其他两种分子的密度都大，所以共价闭合环状 DNA 分子可在密度大的区域形成带而与另两类 DNA 分子分开。

用电镜观察法、中性 pH 沉降法和凝胶电泳法都可以对 DNA 的超螺旋性进行实验测定。在电镜下观察到 DNA 分子超螺旋的程度，如图 2-11 所示。

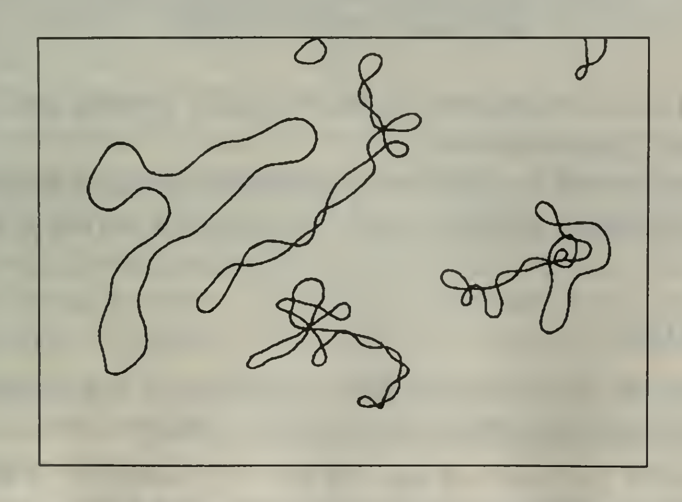


图 2-11　噬菌体 PMZ-DNA 分子的切口环状和超螺旋环状形式

不同构型的 DNA 的沉降系数不同，环状 DNA 分子比线状 DNA 分子的 *S* 值大，共价闭合环状分子和切口环状分子有同样的 *S* 值，而超螺旋分子比其他任何形式的 DNA *S* 值都大。线状分子、非超螺旋环状分子和超螺旋分子的相对 *S* 值分别为：1.00∶1.14∶1.41。

当溴化乙锭插入共价闭合环状 DNA 分子时，在特定浓度下，溴化乙锭插入的 DNA 分子呈负超螺旋，随着溴化乙锭的不断加入，负超螺旋变成松弛环状 DNA，进一步插入溴化乙锭，从而使松弛环状 DNA 变成正超螺旋 DNA 分子，使分子密度更加致密，沉降速度再一次加快（图 2-12）。

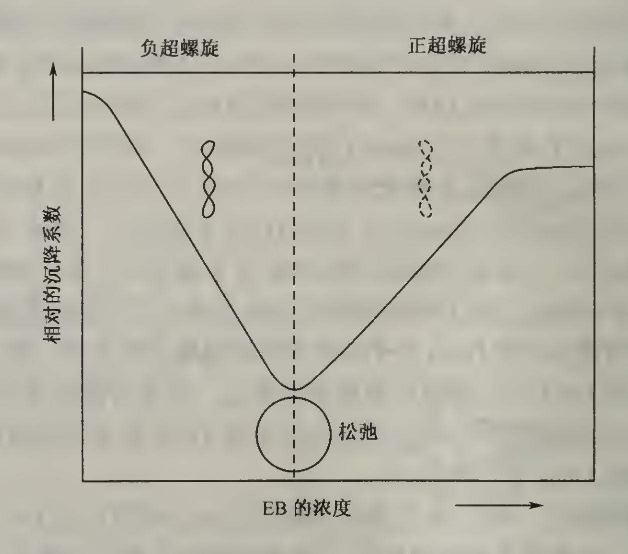


图 2-12　闭合环状 DNA 分子中随着溴化乙锭的插入，
由负超螺旋→松弛环状→正超螺旋

超螺旋 DNA 分子在凝胶电泳中比其他类型的分子移动速度要快，只需相差一个螺旋圈就可在凝胶电泳中分开。

超螺旋的生物学意义主要是：其一，超螺旋形式在 DNA 复制和转录中是必须的（即在基因表达中是必须的）；其二，它可以使巨大的 DNA 分子的体积缩小。

4．左手螺旋 DNA

最早在合成的 CGCGCG 核苷酸序列中发现有旋光性。后来发现只要是嘌呤-嘧啶相间排列的序列在高盐溶液中就可以表现出左旋形式。

在 NaCl 浓度大于 2mol/L 或 $MgCl_2$ 浓度大于 0.7mol/L 时，溶液中的 DNA 的旋光性就会变为左旋，X 射线衍射分析表明这是一种左旋 DNA（Z-DNA），此 DNA 产生的条件是嘌呤-嘧啶的交替出现。

在一条DNA双螺旋中，只有一段是嘌呤-嘧啶交替出现，因此这一段可以采取Z-DNA形式，而其他区段仍为B-DNA形式。形成Z-DNA需要高盐浓度的环境。B-DNA变成Z-DNA是自发的，故认为体内应存在Z-DNA，而用免疫学方法证实了果蝇染色体中确实存在Z-DNA。

用抗体结合法提取的动物病毒SV40基因组中的Z-DNA大部分存在于基因调控区，特别是存在于增强子（enhancer）处。SV40基因组中，交替出现的嘌呤-嘧啶并不多，且均集中在调控区，所以认为Z-DNA的生物学意义是参与基因表达调控。

二、DNA的基本性质

1. 复性

变性DNA溶液用某种方法处理后，使之重新形成天然DNA的过程叫做复性或退火（reannealing）。重新形成的天然DNA叫复性DNA。

（1）复性的必要条件和机制

复性的发生需要两个必要条件：①盐浓度必须高，足以使两链之间磷酸基团上负电荷的排斥力消失，通常用0.15～0.5mol/L的NaCl。②温度必须足够高，以防止链内随机形成氢键。复性的最适温度比 T_m 值低20～25℃。

复性是一个比较慢的过程。实际上重新缠绕成螺旋这一过程并不限制复性速度。变性DNA的两条互补单链之间要形成氢键，必须使配对的碱基处于正确的位置，因为在变性DNA溶液里两条互补链是否处于正确位置是一个随机的过程，所以复性速度与DNA的初浓度有关，关于这一点将在 C_0t 曲线中详细论述。

通过一个假设的含多倍重复序列的分子可以了解复性的分子细节。

1A　　1B　　1C　　1D
''''ATGA''''ATGA''''CCCC''''ATGA''''
''''TACT''''TACT''''GGGG''''TACT''''

假设每条单链含有50 000个碱基且两条链互补，在这样一个DNA分子里，某一短顺序就会提供许多配对位点。lA和1C之间的随机碰撞将产生碱基配对，而两条互补链之间的随机碰撞是无效的。但这种情况是短暂的，因为周围那些短的互补小片段不能稳定配对，从而不能形成稳定的碱基堆积。然而对于1B和1D，这样的顺序一旦形成配对，整个双链DNA分子则可在几分钟内形成。

在变性的DNA溶液中，混合的自由单链碰撞是随机的，每一复性的DNA分子并非都是由原始配对的互补链形成的。随机碰撞也可产生杂交分子。

（2）复性对浓度的依赖关系：C_0t 曲线

复性的起始事件是碰撞，复性的速度依赖于DNA的浓度，同一种DNA，当其浓度高时，互补顺序相碰撞的机会会增加，复性速度也会变快（图2-13）。影

响 DNA 复性的因素很多，除了 DNA 初始浓度外，在同样条件下，顺序复杂的 DNA 比顺序简单的 DNA 复性要慢；DNA 片段大的比片段小的复性慢；温度和离子强度都影响复性速度。

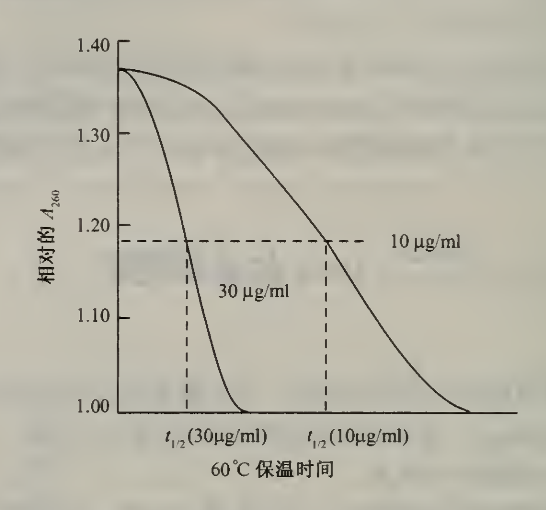


图 2-13　T_7DNA 的复性速度依赖于其浓度

复性过程是一个复杂的多步骤反应过程。复性的第一步是当两条链碰撞时，假如在某个区正确配对，就会形成一个双链核心。核心形成后，互补链就会很快配对从而完成复性过程。因此核心的形成是复性速度的限制因素。

复性过程可用二级反应动力学公式来处理：

$$S + S' = D$$

S 和 S′表示 DNA 互补单链，D 表示双链 DNA。按二级反应动力学公式：

$$\frac{dc}{dt} = -KC^2，积分后得：\frac{c}{C_0} = \frac{1}{1 + C_0 t \cdot K'}$$

C 表示 t 时间的单链浓度，K 和 K' 表示速度常数，t 表示时间，C_0 表示单链 DNA 的初始浓度，C_0t 表示 DNA 单链浓度 C_0 和复性时间 t 的乘积。

在一个指定的实验里，C_0 是已知的，C 可以测定，如以 C/C_0 对 C_0t 作图，可得图 2-13 所示的曲线。在曲线的中点，即 $C/C_0 = 1/2$ 时，$C_0t1/2 = 1/K'$。

在复性实验中，如果我们把初始浓度、温度、离子强度、片段大小都进行控制，那么剩下的唯一可变因素就是顺序的复杂性。实验指出，速度常数 K' 与 DNA 的复杂性 N 成反比：

$$K' = 1/N$$

$C_0t1/2$ 的大小代表了基因组的大小和 DNA 顺序的复杂程度。

用这个方法测定真核 DNA 的 C_0t，发现 C/C_0 对 C_0t 所作的图并不是一条简单的 S 形曲线。例如小鼠 DNA 中 10% 的 DNA 复性速度甚至比已知的最小病毒

DNA的复性还快，说明这10%的DNA是由很多短的重复顺序组成的。对真核生物DNA顺序排列的知识就是从对复性动力学的研究开始获得的。

在DNA的研究工作中，所采用的很多手段都同DNA的变性和复性有关，分子杂交的基础就是DNA的变性和复性。

2. DNA的修饰

DNA分子常常被修饰，这些修饰过程均是在专一性的修饰酶作用下完成的。

DNA中最主要的修饰作用是DNA的甲基化。DNA的甲基化是限制-修饰系统、细胞中DNA复制过程中的错配修复和基因表达调控中的一个重要组成部分。

3. 核酸降解

DNA和RNA在化学方法或酶学方法作用下都可以水解成自由的单核苷酸。

(1) 酸水解

在pH小于1时，DNA和RNA的磷酸二酯键会水解，并且还伴随着碱基和核糖之间的糖苷键的断裂从而产生自由碱基。在pH＝4时，DNA中的嘌呤与核糖之间的糖苷键会断裂，DNA长期暴露在pH＝4的环境中，所有的嘌呤都会脱落。

甲基化试剂可使A、G和C甲基化。甲基化之后，在低pH或中性pH下加热使核糖之间的糖苷键断裂，碱基脱落。也可以利用这种甲基化专一标记某一碱基再加热脱去此甲基化的碱基从而进行序列分析。

(2) 碱水解

当pH大于13时，RNA水解而DNA稳定，DNA在pH＝13时保持稳定，但RNA在pH＝11时全部水解。DNA在pH＝13时不被水解，即其磷酸二酯键不会断裂，但当DNA中的碱基去掉后DNA就很容易被水解。

(3) 酶水解

核酸水解酶多种多样，称为nuclease，根据它们水解的底物不同可分为DNA酶（DNase）和RNA酶（RNase）。许多DNA酶作用于单链DNA或双链DNA，而有些RNA酶只作用于RNA的一个单链末端，一次只除去一个核苷酸，这种酶叫外切核酸酶。外切核酸酶分5′专一性和3′专一性。还有许多核酸酶只在链内发挥作用，这种核酸酶叫内切核酸酶。内切核酸酶一般只识别特定的几个碱基顺序。

核酸酶有多种生物学功能，它已经成为实验室中人们切除核苷酸的有用工具。在RNA顺序测定中的第一步就是使用核酸酶。已分离出的核酸酶的特性见表2-1。

表 2-1　已分离出的核酸酶的性质

核酸酶	底　物*	切割位点	产　　物
Pancreatic ribonuclease	RNA	内切核酸酶；相邻嘧啶	以嘧啶结尾的单核苷酸或寡核苷酸 3′-P 核苷酸
T1 RNase	RNA	内切核酸酶；相邻的鸟嘌呤核苷	以鸟苷 3′-P 结尾的单核苷酸或寡核苷酸
PancreaticDNase I 1	DNA、RNA	内切核酸酶	寡核苷酸
Venom phosphodiesterase	RNA 或 DNA	3′-OH 末端外切核酸酶	5′-P 单核苷酸
Spleen phosphodiesterase	RNA 或 DNA	5′-OH 末端切核酸酶	3′-P 单核苷酸
Micrococcal nuclease	ss DNA	内切核酸酶	3′-P 单核苷酸
E. coli endonuclease I 1	ds DNA	多在 3′→5′-磷酯位点	具有 5′-P 末端的寡核苷酸
Neurospora endonuclease	ss DNA 或 RNA	许多位点	5′-P 单核苷酸
Si nuclease	ss DNA	5′-OH 末端外切核酸酶	5′-P 单核苷酸加上一个末端的二核苷酸
E. coli exonuclease I 1	DNA 单链区	内切核酸酶；任何-ss 位点	5′-P 单核苷酸
E. coli exonuclease V	DNA	3′-OH 末端外切核酸酶	5′-P 单核苷酸
E. coli exonuclease II	ds DNA	3′-OH 末端外切核酸酶	5′-P 单核苷酸
E. coli exonuclease VII	ss DNA	3′-OH 末端外切核酸酶	5′-P 单核苷酸
Phage λ exonuclease	ds DNA	5′-P 末端外切核酸酶	5′-P 单核苷酸

* ds 和 ss 分别代表双链和单链。

(4) S1 内切核酸酶和 S1 图谱

S1 内切核酸酶是目前应用很广的一种酶，S1 酶能切割单链 DNA 或双链 DNA 的单链区，无论这个单链区多么小，即使只有一个核苷酸也能被 S1 酶切开。例如，如果两个 DNA 分子只有一个碱基对不同，这两个分子变性后复性会形成只有一个不配对碱基的杂种分子，S1 酶就可以在这个不配对的位点处切断。这个技术可用于做突变株的物理图谱。一个突变株和野生型 DNA 混合后变性，复性，然后用 S1 酶处理，DNA 则被准确地在突变位点切断，所以可以用 S1 酶确定某一基因突变的位点，因为 S1 酶作用后通过电镜或凝胶电泳分析产生的 DNA 片段的长度可以确定突变的位置。这个技术也叫 S1 图谱（S1 mapping）。另外 S1 酶还可以用于确定转录起始点。提取某一基因的 mRNA 与变性的 DNA 杂交，再用 S1 酶处理，切除未杂交的单链 DNA 部分，加入 pH 为 13 的溶液，使 RNA 水解，保留 DNA，分析 DNA 序列，测知转录起始点。

第三节　RNA

核酸是生物体内重要的高分子化合物，它储存着生物体内全部遗传信息，是基因表达不可缺少的基础物质。除 DNA 外，在生物体内还存在着另一种核酸——RNA。RNA 有许多种类（表 2-2 和表 2-3），它们具有不同的生物功能。

表 2-2 大肠杆菌中 RNA 的含量和性质

种 类	沉降系数	分子质量（U）	核苷酸对数	占细胞总 RNA 的比	代谢性质
RNA	5S	3.5×10^4	120	82%	稳定
	16S	5.5×10^5	1 542		稳定
	23S	1.1×10^6	2 904		稳定
tRNA	4S	$2.5\times10^4\sim3\times10^4$	74～93	16%	稳定
mRNA	4～25S	$2.5\times10^4\sim10^6$	70～3000	2%	稳定
部分其他 RNA	4.5S		107		稳定
	6S		184		稳定
	10S		609		稳定
	4.5S		109		不稳定
	4.5S 前体		130		不稳定

表 2-3 真核生物中 RNA 的存在和分类

位 置	名 称	沉降系数	相对分子质量	核苷酸数	含量*	代谢性质
细胞质	rRNA	5S	3.5×10^4	～120	50%	稳定
		5.8S	5×10^4	～160		稳定
		18S	6.5×10^5	～2000		稳定
		28S	1.7×10^6	～5000		稳定
	tRNA	4S	$2.5\times10^4\sim3\times10^4$	～80	24%	稳定
	mRNA					不稳定
细胞核	tRNA 前体				11%	
	rRNA 前体	45S	4.1×10^6			
		41S	3.1×10^6			
		32S	3.1×10^6			
		20S	9×10^5			
	mRNA 前体		$1\times10^5\sim2\times10^7$			极不稳定
	SnRNA	4～8S		80～260		稳定
线粒体或叶绿体	tRNA	4S（3S）			15%	
	rRNA	类似原核生物或更小				
	mRNA	类似原核生物				

* 为哺乳动物肝细胞中的含量。

1. 成熟 RNA

细胞内含量最多的几类 RNA 分子主要的生物功能是参与蛋白质的生物合

成。第一类是核糖体 RNA（rRNA）。原核生物中有三种 rRNA（5S rRNA，16S rRNA，23S rRNA），真核生物中有四种（一般是 5S rRNA，5.8S rRNA，18S rRNA 和 28S rRNA）。这些 RNA 分子在代谢上十分稳定，是生物体内蛋白质合成的“机器”——核糖体的重要成分。第二类是转移 RNA（tRNA）。生物体内存在着与 20 种氨基酸对应的 tRNA 分子，它们在代谢上也是稳定的，在蛋白质的生物合成过程中起接受、转运和掺入氨基酸的作用。参与蛋白质生物合成的第三类 RNA 分子叫做信使 RNA（mRNA）。它的生物功能是接收 DNA 上的遗传密码即蛋白质中氨基酸排列顺序的信息，并起模板作用合成蛋白质。一般来讲，有一种蛋白质就有一种对应的 mRNA 分子，但有的 mRNA 分子可以翻译出几种蛋白质。mRNA 在代谢上是不稳定的。这几类 RNA 都存在于细胞质中，它们行使生物功能的主要场所在核糖体上。在真核细胞的线粒体和叶绿体里都存在着独立的蛋白质合成系统，因此也存在这几类 RNA 分子。但是，其大小和结构往往和细胞质中的不完全相同。此外，生物体内还存在着多种具有特定生物功能的 RNA 分子，例如，参与启动 DNA 复制的引物 RNA；在某些肿瘤病毒中存在着一些类似 tRNA 的分子是反转录酶的引物，参与 DNA 的合成；在某些细菌中，有的 tRNA 分子在细胞壁合成时可作为甘氨酸的载体；还有一些 RNA 分子是一些酶活性不可缺少的组成成分，例如兔肌 1，4-α 葡聚糖分支酶含有 31 个核苷酸的 2.8S RNA 分子；特异性地作用于 tRNA 前体的 RNase P 中有 77% 的成分是核糖核苷酸。真核细胞的核内还存在着许多小分子 RNA（SnRNA），其大小从 4S 到 8S 不等，大约由 80 到 260 个核苷酸构成，其中有些 SnRNA 在 RNA 成熟过程中起重要作用；另外，还有一些染色质结合的 RNA（chRNA），chRNA 的生物功能尚不清楚。

2. 前体 RNA

绝大多数 RNA 分子都是在细胞核内合成的。核内存在着各种 RNA 的前体，例如 tRNA 前体、rRNA 前体和 mRNA 前体。这些前体 RNA 的分子质量要比对应的成熟 RNA 的大得多，例如哺乳动物肝细胞的 rRNA 的前体是 45S 细胞核 RNA（nRNA），在成熟过程中产生了 41SnRNA、32SnRNA 和 20SnRNA 等中间物，最后相应地成为 28S rRNA、5.8S rRNA 和 18S rRNA。前体 RNA 都是由细胞核 DNA 转录产生的，经过剪切、装配和修饰成为成熟的 tRNA、rRNA 和 mRNA，进入细胞质中各自行使其生物学功能。有人把在细胞核内合成后离开细胞核从而行使生物功能的 RNA 分子称作迁移性 RNA（migrating RNA），而始终存在于细胞核内的叫做非迁移性的 RNA（nonmigrating RNA）。

3. 病毒 RNA

病毒中的遗传物质有 DNA、也有 RNA。绝大多数植物病毒和 RNA 肿瘤病

毒以 RNA 作为遗传物质。一般讲，病毒 RNA（除负链 RNA 病毒外）包括噬菌体 RNA，它既是遗传信息的载体，也具有 mRNA 功能。类病毒（viroid）没有蛋白外壳，只有游离的 RNA 分子。RNA 病毒基因组包括单链 RNA 和双链 RNA 基因组。单链 RNA 病毒基因组又包括“+”链和“-”链 RNA，双链 RNA 病毒基因组包括单分子的和分段的 RNA。RNA 病毒基因组还包括线状、环状 RNA。

这些结构不同的 RNA 基因共同的特点是：在被感染的细胞中，它们必须要有效地复制从而为装配子代病毒提供基因组，又要为合成蛋白质提供 mRNA，由 RNA 病毒执行的这些 RNA 分子的合成是个独特的过程，其细胞中缺少 RNA 指导的 RNA 合成机器。除反转录病毒外，所有 RNA 病毒都编码一种依赖于 RNA 的 RNA 聚合酶，催化新基因组和 mRNA 的合成。

具有“-”链 RNA 和双链 RNA 基因组的病毒粒子必须携带 RNA 聚合酶，因为进入细胞的这些病毒基因组 RNA 不能作为 mRNA 来合成蛋白质，也不能利用细胞机器复制，去除蛋白质的“-”链和双链 RNA 病毒基因组没有感染活性。“+”链 RNA 病毒粒子不携带聚合酶，这些病毒粒子去除蛋白质后有感染活性，因为“+”链 RNA 可以作为 mRNA 在被感染的细胞中通过翻译产生 RNA 聚合酶。

病毒 RNA 聚合酶是数个病毒基因产物组成的常见的多亚基装配体，其中一个亚基催化 RNA 的合成，其他亚基担负其他功能。依赖于 RNA 的 RNA 聚合酶指的是该多亚基装配体而不是具聚合活性的单个亚基。

依赖于 RNA 的 RNA 聚合酶根据其在病毒感染期间的活性，在历史上曾被给于两个不同的名称：复制酶是用于描述该酶拷贝病毒 RNA 以产生大量基因组，而产生 mRNA 的酶称为转录酶。这个命名法指出了执行不同功能的 RNA 合成的酶的真正区别。然而这个术语是不准确的。例如有些 RNA 病毒，基因组的复制和 mRNA 的合成是同一反应。对于双链 RNA 病毒来说，mRNA 的合成产物也可作基因组复制的模板。

转录是指 DNA 中的遗传信息拷贝到 RNA 上的过程，这个术语不适合 RNA 病毒基因组的 mRNA 合成。

4. RNA 的结构

除少数病毒的 RNA 是双链外，RNA 都是单链的。细胞内主要有三种 RNA（rRNA、tRNA 和 mRNA）。这些分子中存在可配对的区域，因此链内可形成双链区，所以像 rRNA 和 tRNA 都有其一定的构型。mRNA 也能形成一定的高级结构，在蛋白质翻译中起调控作用。

5. 核酸的结构分析

(1) DNA 的物理图谱

DNA 的物理图谱就是 DNA 分子的限制性内切核酸酶酶解片段的排列顺序。物理图谱可以用多种方法来测定，其中以部分酶解法和两种不同专一性限制酶交叉酶解法最为简便和常用。部分酶解法是其他各种测定物理图谱的方法的基础。

测定物理图谱的第一步是测定 DNA 完全酶解片段的数目和每个片段的分子量。选择合适浓度的琼脂糖凝胶或聚丙烯酰胺凝胶作为载体进行电泳，通常能把绝大多数酶解片段分开。这些片段的分子质量能从泳动的距离来判断，因为在合适条件下泳动的距离同 DNA 分子质量的对数成反比。如果 DNA 片段是放射性 ^{32}P全标记的，那么就可以从每个片段的放射活性占整个分子放射活性量的百分比推算出此片段分子质量。

测定了完全酶解片段的数目和分子质量之后，接着把 DNA 用限制性内切核酸酶进行部分酶解，测定每个片段的分子质量。从部分酶解片段可以推测哪几个酶解片段是相邻的。例如有一个 $\phi X174$（复制型）*Hin*dII 片段，其分子大小是 1030bp，这个片段用 *Alu* I 可以切成 A、B、C、D 四个片段；分子大小分别是 750bp，150bp，80bp 和 50bp。为了测定这四个片段的排列次序，可以用 *Alu* I 部分酶解，在部分酶解片段中可得到分子大小分别为 810bp，230bp，130bp 等片段。从这个数据立刻可以推断 *Alu* I 酶解的四个片段的排列次序是 150—80—50—750（图 2-14）。当然也可以把每个部分酶解片段从凝胶上洗下来，用同样的酶完全酶解，这样得到的酶解片段应是相邻的片段。

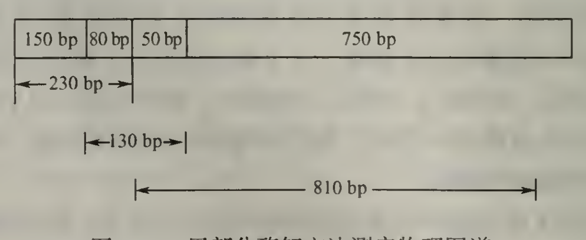


图 2-14　用部分酶解方法测定物理图谱

用两种限制性内切核酸酶交叉酶解也是测定物理图谱常用的方法。当由于某种原因，例如高级结构导致某些部分酶解片段产率特别低时，这就需要第二种酶来帮助。交叉酶解不仅可以得到两种以上限制性酶切图谱，而且可以相互校对，使结果更加可靠。一般先建立产生片段比较少的限制性内切核酸酶的物理图谱，然后在此基础上去建立产生片段比较多的限制性内切核酸酶的物理图谱。

如前所述，部分酶解的方法是建立物理图谱的基本手段，但是部分酶解片段的数目很大，给分离带来困难。现在有一种改进的部分酶解方法，即先将 DNA

分子的一端用放射性同位素标记，部分酶解后电泳分离，那么只有带同位素的片段出现在放射自显影图上，谱线的数目相当于完全酶解的数目，相邻谱线碱基对数目之差就是两个相邻点之间的这段 DNA 片段的大小。例如某个组蛋白基因片段共 5900bp，末端标记之后用 *Hpa* II 部分酶解后得五个片段，其分子大小分别是 5900bp，4210bp，3260bp，2950bp，1020bp，显然最大的部分酶解片段和其次的一个片段核苷酸数之差就是最靠右边的片段分子大小。其余以此类推，五个片段的排列次序是：1020—1930—310—930—1690（图 2-15）。

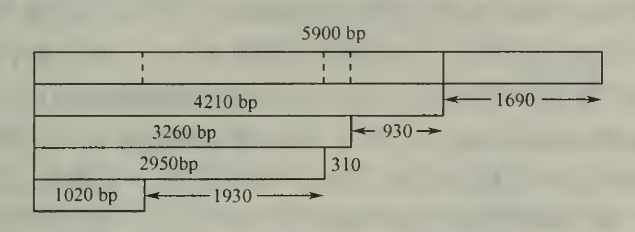


图 2-15　用末端标记测定物理图谱

测定 DNA 物理图谱方法还有很多，在这不一一详述，测定了某一 DNA 分子的物理图谱之后，就可以着手测定每一片段的核苷酸排列顺序了。

（2）测定 DNA 顺序的方法

用于 RNA 和蛋白质顺序测定的基本战略是小片段的重叠。而这个战略用于 DNA 顺序测定时困难就很大，几乎是行不通的。1975 年 Sanger 突破了这个框，提出了一种新的战略。他不是直接测定 DNA 片段中碱基的组成和顺序，而是通过测定片段的分子长度来推测碱基的顺序，用不同的处理方法来确定碱基的性质。其后发展起来的各种 DNA 顺序测定的方法都是以这个原理为基础的。近来这个原理也用到 RNA 顺序测定中，从而发展了很多 RNA 顺序测定的新方法。因此这个原理的提出在核酸的顺序测定中是具有划时代意义的。

这个方法的原理并不复杂。假定有一个 DNA 片段，它的 5′末端用放射性同位素标记，这是片段长度的参考位置。又假定这个 DNA 片段中有两个腺嘌呤，以 5′末端为起点，一个在第六位，另一个在第十位。如果我们设计一种方法，使腺嘌呤相邻的磷酸二酯键断开，而其他碱基相邻的键不受影响，那么将可以得到两个片段，一个是六核苷酸，另一个是十核苷酸，这两个片段可以用电泳方法分开并确定它们的长度。反过来，如果测得片段的长度，我们可以推测，以 5′末端为参考位置，第六位和第十位由腺嘌呤占据着，而其他三种碱基的位置可以用类似的方法推测。显而易见，这个新方法能否成功取决于两点，首先是要有一种方法能把仅仅相差一个核苷酸的同系核苷酸片段都分开。这一点现已办到，例如在 40cm 的聚丙烯酰胺凝胶上电泳就可以把 100～200 的核苷酸全部分开，分辨率达到一个核苷酸水平。其次要有一种手段能制备四种末端碱基专一的片段。

1975 年 Sanger 设计的"加减法"满足了这个要求，此后发展了很多制备片段的方法。DNA 或 RNA 顺序测定方法，原理都是如此。不同方法的区别就在于制备片段的方法不同，按制备片段的方法的不同可以把这些方法分成两大类：一类是 DNA 聚合酶法，简称酶法；另一类是化学降解法，简称化学法。

酶法。以单链 DNA 为模板，加上一个合适引物（一般用限制酶切片段），再加入四种脱氧核糖核苷三磷酸（dNTP），其中一种是同位素标记的。DNA 聚合酶就从引物开始合成一条与模板互补的 DNA 链。对于 DNA 顺序分析最理想的情况是，使合成产物中各种长度的片段都存在。然后从反应混合物中除去四种 dNTP，反应产物分成两部分，一部分用于加法系统，一部分用于减法系统。用于加法系统的产物再分成四小份，每一小份中仅仅将四种 dNTP 之中的一种加入反应系统，比如仅仅加入 dATP，由于 DNA 聚合酶具有 3′→5′外切酶活性，而反应系统中缺少另外三种 dNTP，合成产物就从 3′→5′降解。然而由于存在 dATP，显然，遇到腺嘌呤的位置时降解反应就停止了，因此所有的片段都是以 A 结尾。同理也可以分别制备以 C、G 或 T 结尾的另外三组片段。四组片段在同一块凝胶板中的不同样品槽中同时电泳，从放射自显影图中就可以推断碱基顺序。因为从 A 样品槽中走出的放射性区带代表 A 在片段中的位置，同理也可以定出 C、G 和 T 的位置，因此根据同一凝胶电泳上的四组区带就可以推断碱基顺序。按理只用一个加法系统就足以推断 DNA 顺序，但由于技术上的原因，只用一种方法有时不能得到正确结果，因此又设计了一种减法系统。在减法系统中，也是将上述酶促产物分成四小份，每一小份中只加三种 dNTP，比如缺少 dATP。在这个反应体系中，DNA 聚合酶能够把片段继续合成下去，但是遇到缺少的 dATP 掺入的位置时，合成反应就停下来。这就可以得到一组均以 A 前一个位置结尾的片段，电泳后同样可以测定 A 的位置。图 2-16 是加减法测定 DNA 顺序的原理。

加减法是第一个把片段长度和碱基位置联系起来的新的顺序测定方法，$\phi X174$ DNA 就是用这个方法测定出其全顺序。在 DNA 顺序测定方法的发展过程中，加减法的历史性位置不可低估。但由于技术上的问题，这个方法的实际效果还不够理想。可能由于碱基的特异性和 DNA 的高级结构的影响，聚合反应并非完全同步，反应速度有差异，有些片段可能多些，有的片段可能少些，有时可能导致漏读和重读，有时图谱上还会出现假谱线，同样的碱基连排时图谱上有时只出现一条带。因此用这个方法测定 DNA 片段可能有五十分之一的误差。

对加减法一个最成功的改进是"末端终止法"，其基本反应也是利用 DNA 聚合酶的酶促反应，在反应混合物中有模板、引物、酶和四种 dNTP。除了四种正常的 dNTP 外，另外再加入一定比例的 2′，3′-双脱氧核糖核苷三磷酸，比如 ddATP。因此在 DNA 合成过程中遇到 dATP 掺入的位置就有两种可能的情况发生。如果是 dATP 掺入，合成继续进行，如果 ddATP 掺入，合成反应就不能进

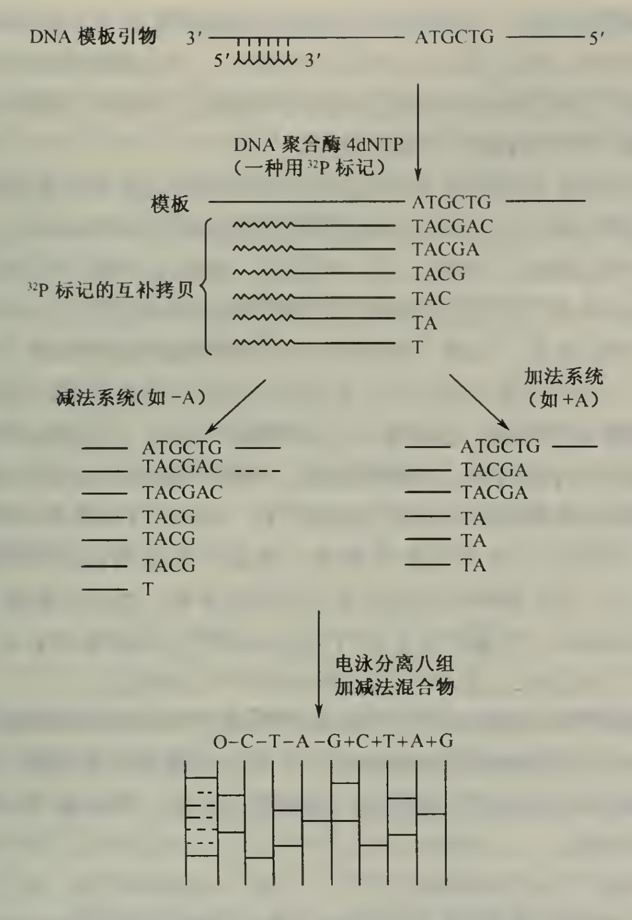


图 2-16　“加减法”测定 DNA 顺序的原理

行了，因为 ddATP 的 3′位置没有羟基。因此可以得到一系列不同长度的以 ddA 结尾的片段。同理，也可以得到以 G、C 或 T 结尾的片段，电泳之后，就可以直接“读出”碱基顺序。这是一个十分方便和精确的方法。1980 年以来，对此技术有进一步发展：把 DNA 重组到单链噬菌体 M13 中，增殖之后，可以获得大量的单链 DNA，而且由于组建的 M13 载体含有通用的引物结合位置，因此可以利用同一种引物。这就简化了单链 DNA 分离和引物制备的步骤，大大推动了末端终止法的发展。以这个方法为主，已于 1982 年完成了 λ 噬菌体 DNA 全顺序（48 502bp）的测定。后来又发展出现了直接从质粒 DNA 模板出发的测序和 PCR 法测序，双链的质粒 DNA 可用碱变性使两条链分开后作测序模板，使 DNA 测序更加灵活、方便。

1982 年洪国藩改进了上述双脱氧与 M13 克隆系统相结合的方法。由于原有方法是随机的，就会不可避免地出现测定的多次重复。引进后的方法叫做

"DNA 顺序连续测定法"。连续法的关键是将克隆的外源 DNA 的一端固定起来，而另一端用没有专一性的 DNase I 将其逐渐缩短，然后找到长度约有 200bp 差异的克隆，实验设计使缩短的一端发生在共同引物区，这样对克隆的顺序进行逐个测定，就实现了对长链 DNA 的连续测定。

化学法。Gilbert 及其同事 Maxam 利用某些化学试剂专一地修饰碱基，使相应的糖苷键变得不稳定，这样可以得到以特定碱基结尾的四组片段。他们所选用的化学试剂一种是硫酸二甲酯，另一种是肼。硫酸二甲酯可以使鸟嘌呤的 N-7 位和腺嘌呤的 N-3 位甲基化，甲基化后的糖苷键是不稳定的，在中性 pH 时加热就会使甲基化嘌呤脱落，脱落了嘌呤的 DNA 在碱性溶液加热时，磷酸二酯键就会断开，这样一组片段进行电泳后，在放射自显影图上有一深一浅两组谱线，深的是由于 G 降解而产生的，浅的是由于 A 降解产生的。G 之所以比 A 深是因为 G 甲基化的速度比 A 快五倍。如果将硫酸二甲酯甲基化的 DNA 片段先用稀酸处理，则脱嘌呤反应主要是在 A 的位置上进行，因为在这种条件下甲基化的 A 比甲基化的 G 更不稳定。然后再断开磷酸二酯键，电泳图谱上深的谱线代表 A，浅的谱线代表 G。对于嘧啶可以选用水合肼部分肼解，然后用哌啶催化 β 消除反应，使磷酸基团脱去，此条件对 C 和 T 没有选择性，当肼解时存在 2mol/L NaCl 即可抑制肼同 T 的反应，这样产生的谱线相当于 C 的位置。

在对两种碱基专一的化学反应中，酸处理通过质子化作用使嘌呤糖苷键不稳定，碱处理优先打开腺嘌呤和胞嘧啶环，最后均可被哌啶所消除，使 DNA 链在 G 和 A（酸处理）或 A 和 C（碱处理）处断裂。此外，甲烯蓝、四氧化锇也曾被用于 DNA 顺序分析。

表 2-5 列出的三组化学降解反应中，以第 II 组效果较好。若在第 II 组再加上 A>C 反应（见第 III 组的 9），共五个反应，能得到更为满意的结果。

化学降解的精细程度在理论上应达到 100%，但实际上测定中并非如此理想。如果用厚的聚丙烯酰胺凝胶板（1.5mm），精确率可能只有 85%；如果从两个方向测定则可提高到 99%；如果用薄的胶板（0.3mm），一次测定精确度则可达到 99%。

DNA 顺序测定方法主要是化学降解法和末端终止法，也有一些其他方法。但基本上还是模仿化学降解法和末端终止法。事实上除了这两种方法，其他方法很少有人采用。化学降解法和末端终止法各有利弊。对于单链 DNA 而言，末端终止法特别合适，但化学降解法由于条件简单，很多实验室也很乐意采用。近年来由于将末端终止法与 M13 克隆系统相结合，已经把 DNA 顺序分析推向一个新阶段。由于方法上的突破，DNA 顺序测定已成为分子生物学中发展最快的分支之一，短短几年内就完成了不少 DNA 的全序列分析。第一个完成的全序列分析是 $\phi X174$ DNA（5386bp，1977 年）。接着测定了 SV40 DNA（5226bp，1977 年）、G4 DNA（5577bp，1978 年）和 fd DNA（6408bp，1978 年）。1981 年报道

了人和牛线粒体 DNA（16 569bp 和 16 338bp）的结构。1982 年发表了 λ 噬菌体 DNA（48 502bp）的结构。

前些年，人们还千方百计设法把 DNA 转录成 RNA，用测定 RNA 的方法来反推 DNA 的顺序。现在，DNA 顺序测定的方法达到的水平之高足以通过对 DNA 进行测序从而推出 RNA 的序列。已经完成的一些大的 RNA 的测序如卵白蛋白的 mRNA、16S rRNA 等都是将其反转录成 DNA 后测定顺序的。总的来说，RNA 顺序测定进展不如 DNA。

如果对 DNA 顺序测定作一乐观估计，可以设想现在所有的 DNA 顺序原则上都能测定，只是工作量大小的问题。

（3）RNA 的结构分析

主要有两种方法用于 RNA 顺序测定。一种是片段重叠法，另一种是直读法。如前所述，由于 DNA 顺序测定的方法简单易行，所以现在一般采用反转录酶将待测的 RNA 反转录成 DNA，再通过对 DNA 的顺序测定来推断 RNA 的结构，RNA 的顺序测定方法不再详述。

第四节　蛋　白　质

所有的蛋白质都是氨基酸的多聚体。从蛋白质的氨基酸顺序测定中发现，每种蛋白质都具有特定的三维结构，与一般只具有双螺旋结构的 DNA 相比，对蛋白质的研究更为复杂。但从另一方面来讲，由于蛋白质的结构千差万别，使得从细胞里制备某一种蛋白质成为一个十分容易研究的课题。

一、蛋白质的结合位点和多亚基蛋白质

细胞中的每一种蛋白质都执行一种特殊的功能，主要是结构的功能、酶学的功能和调节的功能。执行结构功能的蛋白质通常是纤维状的，而执行另外两种功能的蛋白质是球状的，执行酶学和调节功能的蛋白质催化两种蛋白质分子之间或者细胞内多种分子之间的高度专一反应，这些反应发生在蛋白质分子表面的氨基酸侧链之间。

由于多肽链的折叠，侧链会产生一个带有电荷和疏水基团的独特的表面构型，通常存在一个表面构型大小为 5～10Å 的小区域（这个小区域叫蛋白质的结合位点），它与蛋白质和其他分子之间的相互作用有关。一个酶的结合位点是它的催化活性区，即活性位点，由多肽链上相距较远的氨基酸侧链通过链内相互作用形成一个结合位点（这足以使人们明白为什么在突变株中，一个氨基酸被另一个氨基酸的替换如果发生在结合位点上有更明显的效应，甚至可以通过改变链的其他部分的折叠从而破坏结合位点）。

一个小分子对结合位点的结合可以有许多方式，图 2-17 是其中的几种。

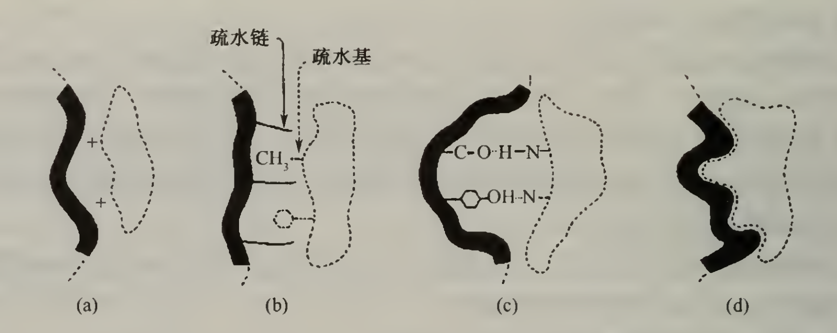


图 2-17　小分子与蛋白质分子结合的四种类型
(a) 电荷；(b) 疏水作用；(c) 氢键；(d) 范德华力。

a. 蛋白质上的两个带电基团可以以某种方式形成一定的空间结构与小分子上其他两个带电基团准确地配对结合。

b. 蛋白质中的侧链疏水簇可提供一个具有非极性的表面，使一个具有很大非极性基团的分子吸附在上面。

c. 蛋白质中的氢键基团很容易与小分子的互补基团形成氢键。

d. 蛋白质表面以某种排列形成一种电子构型，使这种蛋白质分子和将要结合的分子之间具有正确的形状，从而使本来很弱的范德华力能以这种方式产生很紧密的结合，这就叫做空间配合（steric fit)。

通常这几种结合模式是同时存在的。例如，在绝大多数情况下，酶的活性位点的形状和酶作用底物分子的形状接近互补，这样使存在的范德华引力相当强，并且具有在结合位点排斥那些不合适的分子的效应。由一种或者两种这样的吸引力相互作用形成的结合通常是稳定的，这些吸引力即离子键、疏水相互作用或氢键。

一个小分子或大分子结合到一个蛋白质的结合位点可能会引起该蛋白质形状的微小改变，因为结合常常改变决定一种蛋白质形状的那些作用力的强度。形状的改变为蛋白质提供了两种重要功能，其一，一个酶的形状的改变能够改进酶的活性位点和它作用的底物分子之间的接触，酶的这种性质叫做诱导适合（induced fit)；其二，通过形状的调整可调节蛋白质分子的生物活性。如果一个酶有两个结合位点，一个调节分子可以结合到一个位点，由此而影响另外一个位点。影响的方式可能是阻止或促进活性位点的结合。

一种蛋白质分子的表面可以结合另一种蛋白质分子。例如，蛋白水解酶类(proteinase）和蛋白质聚集在一起形成一个复杂的结构。又如，膜和病毒外壳尤其是病毒外壳以相同部件自我聚集形成一个多亚基的复合物。

一个多肽链通过折叠后，非极性侧链往往在内部，即与水是分开的。因而对

于一条多肽链来说，以这种方式折叠后，其非极性基团可能完全在内部，这样将使得非极性氨基酸在此成簇而力求最大限度地减少它与水的接触。一个蛋白质的大疏水区与另一蛋白质的大疏水区配位从而进一步减少它与水的接触。同样，如果一蛋白质分子具有距离较远的几个疏水区，这样的几个相同的分子相互作用形成的结构就能更有效地减少其与水的接触。这种蛋白质是同亚基蛋白质。同亚基蛋白事实上是很普遍的现象，一般以 2 个、3 个、4 个或 6 个亚基组成的同亚基蛋白最为常见（多亚基蛋白也可由不同的亚基组成，这一点不细讨论）。

蛋白质为什么要采取多亚基形式呢？理由是，这种形式可以有效地配置蛋白质的疏水基团。另一个可能会更正确的观点是：许多蛋白质以逐渐形成疏水区的方式折叠，同时亚基将按这个次序装配。这样的进化是有价值的，因为多亚基体系有以下优点：

1）亚基对 DNA 的利用来说是一种经济的方法。合成一个 6000 氨基酸的蛋白质需要 18 000 个碱基对，也就是说相当于 11 个典型细菌 0.04％的编码容量。如果这个蛋白质是由 6 个相同的亚基装配而成的，那么仅 3000 个碱基对就够了。一个多亚基蛋白可以像一个单一大分子蛋白一样形成结合位点。正像某一特定的多肽链上相距较远的氨基酸在一起形成一个结合位点一样，结合位点也可以由不同亚基上的氨基酸连在一起而形成。

2）采取多亚基形式可以减少蛋白质合成过程中随机的错误对蛋白质活性的影响。如果在蛋白质合成期间一个 6000 氨基酸的蛋白质中插入了一个错误的氨基酸，那么这个 6000 氨基酸的蛋白质的最佳拷贝也是无功能的（当然是假设，并非总是这样，只是举例）。如果此蛋白质由 6 个亚基组成，每个亚基含 1000 个氨基酸，6 个亚基中有 1 个有错误，那么以亚基形式组成蛋白质的这种安排可以从两方面保护蛋白质的功能。第一，这个蛋白质可能在某种程度上有功能，这样一个过剩体系形成无功能的可能性比一个非过剩体系要低。第二，一个缺陷的亚基就像一个具有不同表面结构的亚基一样，将完全掺入到多亚基蛋白质中，因而通常可形成一个正常的蛋白质。所以，蛋白质如以完全相同的多肽链组成的话，将完全消除错误的影响。

3）多亚基蛋白质的活性开关能够非常有效地和非常迅速地被打开和被关闭。这种现象在酶的活性调节中有很多例子。一个酶通常只对直接的一步反应有活性而对下一步没有活性。例如，在肌肉的收缩和松弛循环中，在许多情况下，蛋白质通过亚基的解离抑制活性，或通过聚合而恢复活性。这个循环方式的完成过程如下：一个信号分子或效应分子结合到一个亚基上使其形状发生很轻微的改变，此亚基从多聚体上掉下来，游离出来后其形状随之改变，从而失去了与效应分子结合的能力，失去效应分子后，亚基又恢复其原来的形状，并且又能重新聚集形成一个有活性的蛋白质。这类开关是很迅速的，使活性降低到零的水平只需消耗很少能量。

多亚基蛋白质除了目前讨论的相同亚基的组成形式外，不同的多肽链也可以聚集形成由不同亚基组成的蛋白质。实际上这是相当普遍的现象。例如，血红蛋白——血液中的氧的携带者，由 4 个亚基组成（四种亚基为两种类型）。类似的还有催化 RNA 合成的 RNA 聚合酶，是由 5 个亚基组成，其中 4 个是不同的。而在 *E. coli* 中合成 DNA 的 DNA 聚合酶 III，是由 10 个不同的亚基组成的。

一个小蛋白质含有亚基是罕见的，由于多数的蛋白质都由亚基组成，而且多亚基蛋白质又有众多优点，那么是否大蛋白比小蛋白好，当然这不能绝对而论。亚基的增多可能刚好是一个进化中的意外事件。例如，一个突变能使一个多肽链发生自身聚集，这样反而能产生一个活力位点从而催化一个新的反应，如果这个酶对发生突变的细胞是有价值的话，那么这个生物宁可选择这个结构简单的酶而不必采取多亚基酶了。大蛋白质有两方面的好处，第一，小蛋白很难形成一个与很大很复杂的分子结合的结合位点，而大蛋白质很容易办到。第二种情况是当一种结合反应可被调节，这时结合蛋白至少要有两个结合位点，一个结合被结合的分子，另一个用以结合调控因子，小分子质量的蛋白质几乎不可能有两个或两个以上的结合位点。

二、蛋白质活性的调节

许多蛋白质的活性随着细胞内和细胞外的条件而变化。在此我们将讨论蛋白质的活性被调节的几种方式。

1. 终产物的抑制作用

如果一个酶催化 A→B 的反应，而当产物 B 由外源提供时，这个酶没有理由再继续催化这个反应。事实上，如果 A 有别的用处，它将有效地关闭这个酶的活性，要完成这一步，一个简单的办法是使这个酶具有两个结合位点，一个结合 A，另一个结合 B。如果结合了 B，这个酶就改变构象而破坏结合 A 的结合位点。B 以这种方法阻止酶执行 A→B 的转换，这是终产物抑制作用的一个例子。

2. 多亚基蛋白质的调节——变构

许多酶为了节约遗传信息而含有几个亚基，这些酶的活性也是可以被调节的。然而，底物结合的位点和抑制物（可能是反应产物）结合的位点是在不同的亚基上，如果抑制物结合，就阻止了底物的结合。一个亚基上的结合位点被占据的这个信息必须以某种方式转递到其他亚基，这可能是以下述方式通过亚基接触而介导的：抑制物分子的结合改变这个亚基的形状，形状改变将影响这个亚基与其他亚基相互作用的结合位点［图 2-18（b）］，如果这个亚基仍保持接触，那么所有毗邻的亚基首先将在各亚基的相互作用位点处改变形状，这就会依次改变其

他亚基的结合位点，蛋白质经受这样的修饰，叫做蛋白质的变构。

一个抑制物可以降低一个蛋白质的结合活性，相反地，一个无活性的蛋白质也能被激活。换句话说，一个亚基上的底物结合位点对于底物的结合本来不是十分正确的，但当一个激活物分子被结合到另一个亚基的结合位点时，它就可能获得正确的构象［图 2-18（a)］，这样，一些蛋白质经历变构而被激活，而另一些蛋白质经历变构而被抑制。

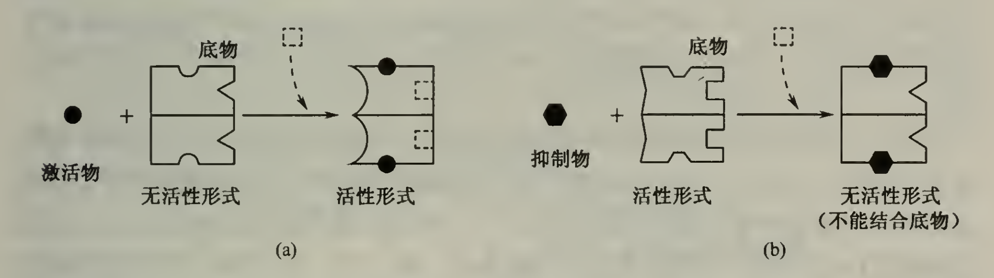


图 2-18　变构的一个例子

（a）激活；（b）抑制。

正是这两种情况可以解释开→关（抑制）和关→开（激活）过程。而许多多亚基蛋白质则根据更微妙的变化——即它们的结合活性是由一个调节因子来调节。

以上我们涉及的是第二个分子如何影响底物的结合。而底物分子本身就可以调节一个蛋白质的结合活性的例子更为普遍。例如，催化 A→B 转变的一个酶，当 A 的浓度低时，其活性必定很弱，而当 A 的浓度高时，酶活性就高。这个过程靠一个单亚基蛋白质是很难完成的。一个多亚基蛋白的每一个亚基可以通过较低的吸引力结合 A，但是一旦 A 结合到一个亚基上就可能引起构象改变而增加其他亚基上的结合位点对 A 的吸引力。如果存在的 A 很少，那么起始结合的频率就很低，因而将只有少数结合位点被 A 占据，这样 A→B 的转变就很少发生。如果 A 的浓度高，第一个结合位点被占据从而增加其他位点的吸引力，这些位点将迅速被 A 充满，而当一个 A 转变成 B 后，这个 A 的结合位点立刻重新被占据，所以 A 浓度转换速率也很高。血红蛋白对氧的结合能力随氧的浓度增加而增加就是一个重要的例子。

3. 变构调节的两种机制

变构的变化可以几种方式发生，曾提出两种重要的模型来描述这个机制，这两种模型叫做对称（或协同）模型和连续模型。两种模型假设如下：①假设每个亚基具有形成两种形式的能力，T 和 R 型，它们分别以低的和高的吸引力与 A 结合。②当 A 牢固地结合到一个亚基上，这个亚基必须是 R 型。

这些模型在几种方式上有区别：

第一，协同模型假设 T 和 R 两种形式的存在最初是平衡的。只有 R 型亚基与 A 牢固结合（T 型亚基与 A 的结合很少或很不牢固），并且与一个 R 型亚基的结合可引起同一蛋白质中的其他所有亚基从 T 型变成 R 型，这样，平衡明显地向 R 移动。相反，在连续模型里，假设 A 首先与 T 型亚基结合（低吸引力的结合），并且这种结合以某种方式诱导这个 T 型（不是其他的 T 型）转变到 R 型（可强有力地结合 A）。注意在协同模型中底物分子与一个亚基的结合可间接影响其他亚基的形式，而在连续模型中，每个亚基都只能直接被影响而不能间接被影响。

第二，结合位点没有被占满的那些变构蛋白可能存在的形式的数目对于两种模型来说是不一样的：在协同模型中认为亚基排列的对称性是迫使大分子聚集稳定化的一个基本因素。在协同模型中，一个 R 型亚基可以稳定地与另一个 R 型亚基相互作用，两个 T 型亚基也可以聚集在一起，但一个 R 和一个 T 则不能形成稳定的一对。只要这些亚基保持接触，一个 R 型亚基起的稳定作用就能引起所有保持着 T 型的亚基转变成 R 型亚基，即蛋白质聚集保留对称性。例如，对一个四聚体来说，协同模型只允许 TTTT 和 RRRR 这两种形式的复合物，而 TTTR、TTRR、TRRR 是不存在的。在连续模型中，上述观点就不适合了，一般认为在连续模型中从 TTTT 变成 RRRR 经过以下几个步骤：TTTT→TTTR→TTRR→TRRR…→RRRR。在这个过程中，每一步的转变都需要另一个分子的结合（图 2-19）。

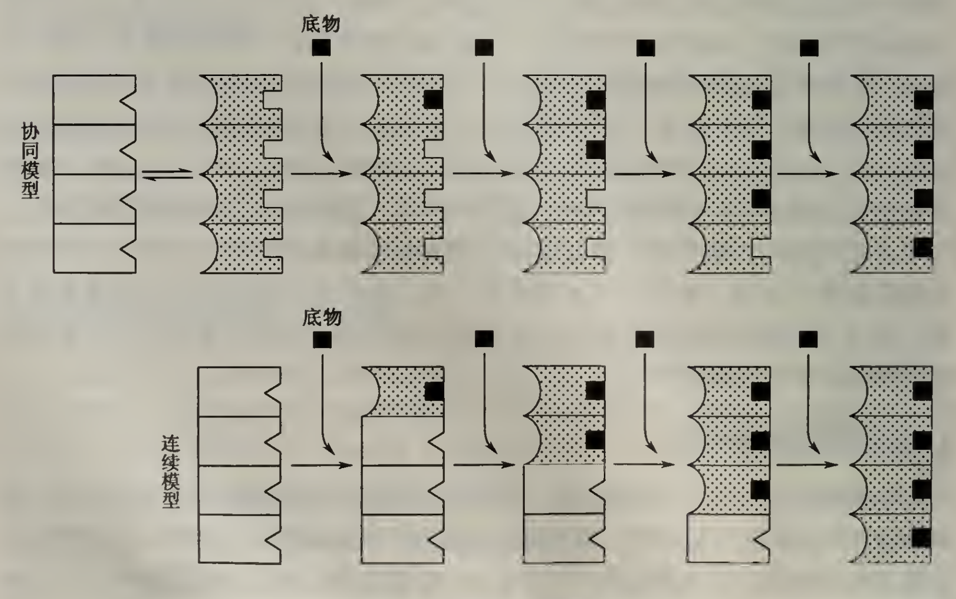


图 2-19　变构的主要模型

协同模型可以通过一个简单的数学形式来描绘，即被结合的 A 与自由 A 的浓度有关。关于连续模型中的平衡则更为复杂。

实验表明，一些大分子可以用协同模型给予最好的阐明，而对另一些大分子，用连续模型来解释要好些。但是有许多分子用两种模型都不能很好地解释它们，这样的一些分子以两种以上的状态存在，对于这些分子的性质需要更复杂的模型来解释。

4. 一种变构酶：天门冬酰胺转氨甲酰酶

变构蛋白研究得最清楚的例子是天门冬酰胺转氨甲酰酶（ATCase），它具有结合激活物和结合抑制物的两个结合位点，它可以催化谷酰胺合成 RNA 前体物——胞嘧啶核苷三磷酸（CTP）。反应顺序是：

谷酰胺 + CO_2 + ATP→氨基甲酰磷酸盐

↓ 天门冬酰胺转氨甲酰酶

天冬氨酸盐 →

CTP←UTP←UMP←甲酰胺天门冬氨酸

这个酶分为两类，共 12 个亚基组成。其中 6 个是催化亚基（C，catalytic subunit）形成两个三聚体。其余 6 个是调节亚基（R，regulatory subunit），调节亚基形成三个二聚体。（注：这里的 R 指调节亚基，不是前面讨论的 R 型。）每个 R 二聚体与两个 C 亚基接触，但接触的两个 C 亚基分别在不同的三聚体中（图 2-20），实际上这些亚基的排列已经用 X 射线衍射研究进行证明（如图 2-20），ATCase 比只由一种类型的多肽组成的四聚体要完善得多。这种酶的行为不能用协同模型或连续模型来解释，尽管它们的变构原理都是一样的。

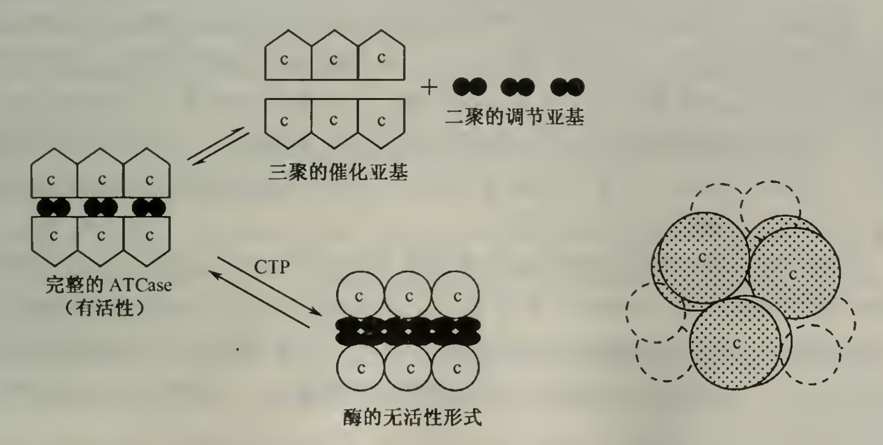


图 2-20 天门冬酰胺转氨甲酰酶的结构与亚基排列

作为一个变构酶，ATCase 以下述方式起作用，此酶被 RNA 的前体物 ATP

所激活，被 CTP 所抑制。CTP 不是中间产物而是终产物（图 2-21）。在实验室条件下 ATCase 很易解离释放出两个 C 三聚体和三个 R 二聚体，并且可将它们分离提纯，提纯后发现底物天门冬氨酸只结合在三聚体上（每个 C 亚基上只结合一个天门冬氨酸）。ATP 和 CTP 分别只与二聚体结合（每个 R 亚基上有一个 ATP 结合位点和一个 CTP 结合位点）。一个 C 三聚体单独就可合成氨甲酰天门冬氨酸，但其活性不受 ATP 和 CTP 浓度的影响，所以酶的激活和抑制需要 R 二聚体和 C 三聚体的接触。同样，发现当 CTP 与之结合时，R 亚基经历了变构的变化，这样的亚基如果与 C 三聚体保持聚集，C 亚基的形状也将发生改变，这与变构效应的原理完全一致。

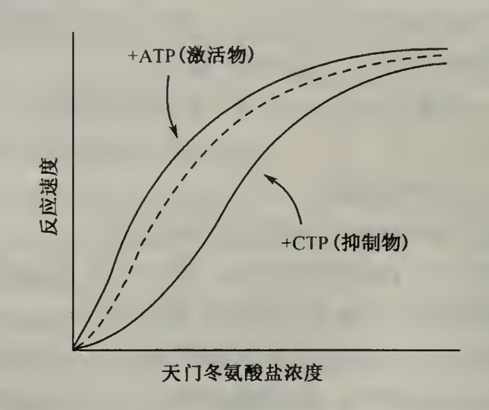


图 2-21　天门冬酰胺转氨甲酰酶的变构效应

5. 真核细胞中蛋白质的磷酸化是导致变构的共同方式

在细菌中，例如在大肠杆菌中的蛋白质的活性调节是通过细胞中无数的小分子结合到特定蛋白质上引起蛋白质的变构，从而控制蛋白质的活性，以这种方式调节的许多蛋白质是催化分解代谢反应的酶，另一些是传递基因开关信号的蛋白质。一些细菌将磷酸基团共价结合到蛋白质的氨基酸侧链上，因为每个磷酸基团带两个负电荷，因此可以吸引多肽链上带正电荷的侧链簇而改变蛋白质的构型，从而以这种方式调控蛋白质的活性。

真核细胞中蛋白质活性调节的主要战略是蛋白质的可逆磷酸化。一个典型的哺乳动物细胞中约有 10 000 种蛋白质，其中 10% 以上被认为是被磷酸化了的。蛋白激酶将 ATP 上的磷酸基团转移到蛋白质上，或者磷酸酯酶将磷酸基团从蛋白质上除去。真核细胞中含有许多蛋白激酶和磷酸酯酶，这些酶中大多在细胞内信号传导中起核心作用。

cdk（cyclin-dependent protein kinase）蛋白激酶能作为一个微凿子（microchip）发挥作用。在真核细胞中有数百种蛋白激酶组成复杂的信号途径网络，该途径协调细胞的活动，驱动细胞周期，并把信号从细胞的外环境中传递到细胞

内，许多被传递的信号需要被整合和放大，单个蛋白激酶（和其他信号蛋白）作为加工手段，或者“微凿子”在整合过程中起作用。

胞环素（cyclin）依赖的蛋白激酶是这种加工手段的一个具代表性的例子。这一类激酶是真核细胞中细胞分裂控制的核心成分。在脊椎动物细胞中，各个cdk酶在整个细胞分裂周期的不同时期按程序连续地打开或关闭，当它打开时，通过它们在蛋白质上磷酸化的效果而多方面影响细胞行为。

一个cdk蛋白，只有当它被结合到被称为胞环素（cyclin）的第二种蛋白时，才具有蛋白激酶活性。cdk的激活有三种不同的方式，胞环素的结合只是其中之一，通过磷酸化和去磷酸化，cdk被激活从而驱动一套特殊的细胞成分——胞环素、蛋白激酶和蛋白酯酶，使之处于合适的活性状态。例如某些胞环素的总量在细胞周期中逐渐增加，直到在细胞周期中某一特定点突然被摧毁。一个胞环素被靶蛋白酶突然摧毁将立即关闭cdk酶活性，这是控制细胞内诸如有丝分裂等事件的一个重要方法。

cdk的三维结构暗示了这种酶调控的分子解释。cdk蛋白本身有ATP结合位点并且是扭曲的，一个大约20个氨基酸的易弯曲的环可封闭蛋白质底物进入到活性位点的通道。胞环素的结合即能消除扭曲又能允许激活的磷酸基团加到易弯曲的环的尖端，于是一般认为该磷酸被吸引到一个带正电荷氨基酸的袋子中，将环拉低以便允许底物进入到活性位点（图2-22）。胞环素的结合也允许抑制性磷酸加入从而干扰ATP位点，这就保持cdk蛋白处于无活性状态。当一种磷酸酯酶除去抑制磷酸时，这种激酶最终被激活（图2-23）。

结合和水解GTP的一类蛋白质是普遍存在的细胞调控因子。我们已经讨论了细胞如何利用磷酸化和去磷酸化来控制该蛋白质的活性。蛋白激酶将ATP上的磷酸基团直接转移到蛋白质的氨基酸侧链上。真核细胞也利用其他方法通过磷酸化或去磷酸化来控制蛋白质活性。在这种情况中，磷酸基团不是直接加到蛋白质上，而是GTP的一部分，GTP与蛋白质紧密结合。结合GTP的蛋白质是有活性的，当GTP被其结合的蛋白质水解成GDP时，就会丢失一个磷酸基团，从而使结合GDP的蛋白质无活性。

GTP结合蛋白（也称GTP酶，因为它水解GTP）组成一个蛋白质大家族，这类蛋白质都有类似的GTP结合球状区域。当它结合的GTP被水解成GDP时，此区域将经历变构的变化而使蛋白质无活性。

一种小的GTP结合蛋白称Ras蛋白，Ras蛋白在细胞信号转导中起决定性的作用。结合有GTP的Ras蛋白是有活性的，它能刺激细胞内蛋白质级联式地磷酸化。然而蛋白质在大部分时间中处于结合GDP的无活性状态。只有当细胞外信号诸如结合在细胞质膜受体上的生长因子作出应答，它结合的GDP变为GTP时才被激活。如此Ras蛋白起一个开关作用，它的活性仅取决于在其结合的GDP上是否加上一个磷酸。这和cdk蛋白活性控制的方式一样。其他一些蛋

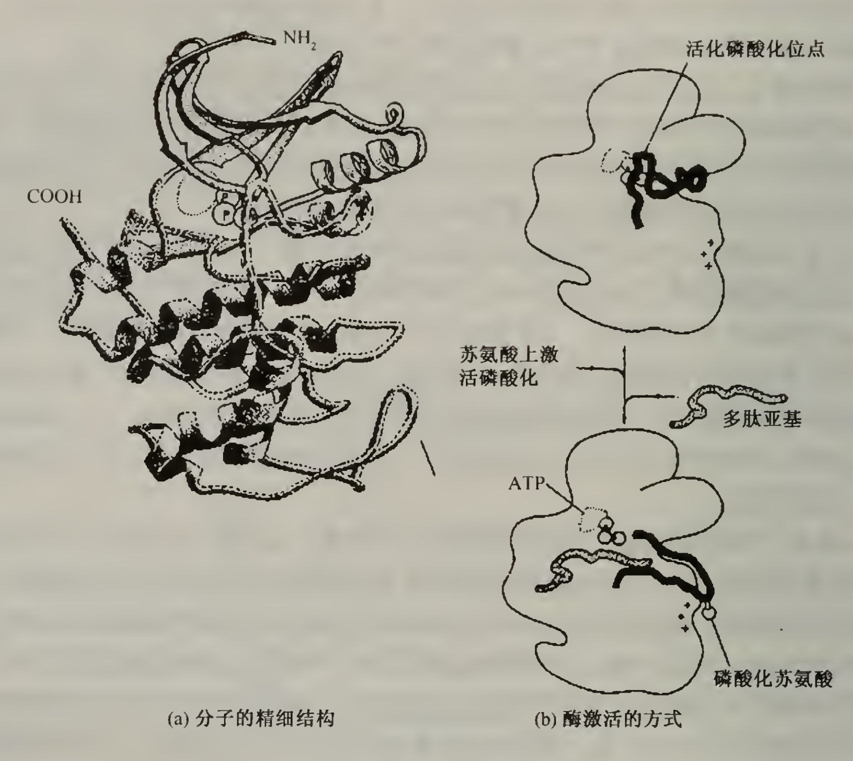


图 2-22　cdk 酶的三维结构

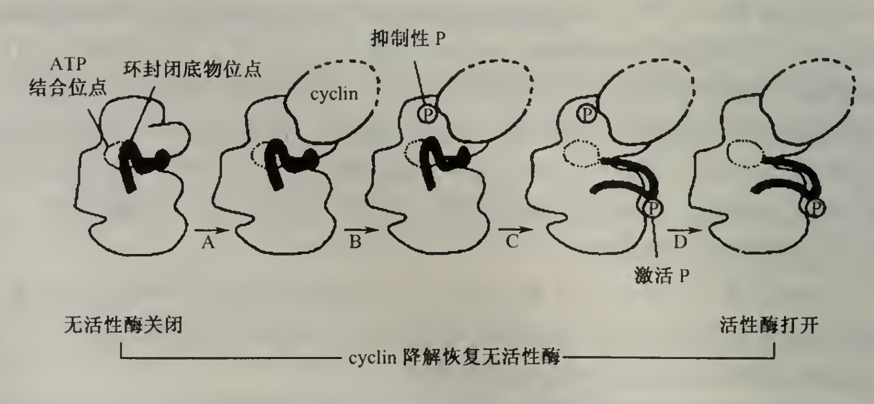


图 2-23　cdk 激活的精细模型

白质也是通过结合 GDP 或 GTP，来控制 GTP 结合蛋白的活性。

正像 cdk 蛋白的活性被胞环素、蛋白激酶和蛋白磷酸酯酶控制一样，Ras 蛋白和其他 GTP 结合蛋白的活性也是取决于结合 GDP 还是结合 GTP。Ras 蛋白被 GTP 酶激活蛋白（GTPase-activating protein 或 GAP）灭活，GAP 结合到 Ras 蛋

白上并诱导它对结合在蛋白质上的 GTP 水解，使 GTP 变为 GDP，且牢固地结合在上面。因此该 Ras 蛋白将处于结合有 GDP 的无活性状态，直到它遇到一个鸟苷酸释放蛋白（guanine nucleotide releasing protein，GNRP），GNRP 能结合到 GDP-Ras 蛋白上，引起 GDP 释放。因为空出的核苷酸结合位点立刻被 GTP 分子占满（在细胞中 GTP 比 GDP 多很多），所以 GNRP 是间接地将 GTP 酶水解除去的磷酸又加回去从而激活 Ras 蛋白。因此从某种意义上讲，GAP 和 GNRP 的作用分别和蛋白磷酸酯酶和蛋白激酶的作用类似。

EF-Tu 蛋白中的变构转变说明了一个小的变动如何产生大的移动。

Ras 蛋白是单体调控因子 GTP 酶类的一个成员，这个家族中每个成员均有大约 20 个氨基酸的 GTP 结合域（domain）。在进化进程中，这种区域也可以加入到其他蛋白质区域而形成一个 GTP 结合蛋白大家族，该家族中的成员包括受体-相关三聚 G 蛋白、空泡和细胞内小室之间的交通调控蛋白，结合到 RNA 上的这类蛋白质对核糖体上的蛋白质合成是必需的。这类蛋白都是通过 Ras 类区域中的 GTP 的水解而引起蛋白质构型改变来调控重要的生物活性。

EF-Tu 蛋白为这个家族中蛋白如何工作提供了一个范例。EF-Tu 是细菌细胞中丰富的分子，作为蛋白质生物合成的延伸因子，装载每个胺酰 tRNA 分子进入核糖体。tRNA 分子与结合有 GTP 形式的 EF-Tu 形成紧密复合物。在这个复合物中，tRNA 上携带的氨基酸被遮蔽，当随着结合到 ET-Tu 上的 GTP 水解，tRNA 被释放时，核糖体被除去遮蔽并暴露出氨基酸，这对蛋白质的合成是必需的。

已经用 X 射线衍射测定了结合 GTP 和结合 GDP 形式的 EF-Tu 的三维结构，研究显示出了 tRNA 的去遮蔽反应是如何发生的。随着 GTP→GDP + Pi 的反应，无机磷酸基团释放，引起在 GTP 结合位点的几十分之一毫微米的滑动，这正如在 Ras 蛋白中进行的那样。这个相当于一个氢原子直径几倍的小小移动，可引起一个构型改变并沿着该蛋白质中 Ras 类似域中称作开关螺旋的一个十字形 α 螺旋传递。这个开关螺旋似乎可以作为一个门闩黏着到这个分子的其他区域中的特殊位点上，使蛋白质处于关闭构型。GTP 水解引起开关螺旋变构使两个区域分开，分开的距离大约 4nm，以便 tRNA 释放（图 2-24）。

从这个例子中人们可以看到细胞是如何从一个小的蛋白质区域上发生的简单的化学变化发展成具有复杂功能的大蛋白质。

水解 ATP 的蛋白质在细胞中做机械功。像用于调控化学反应一样，变构改变形状也可用于调控细胞中的有次序的运动。例如，假定需要一个蛋白质沿着像 DNA 这样的分子一样的细线“步行”，那么变构蛋白可以通过一系列的构型改变也许能完成这样的工作。

如何使一系列的构型改变呈单一方向，使一个完整的循环过程往一个方向进行，使得任何一步都不可逆。这些工作所用的机制正如通过 GTP 水解驱使一个

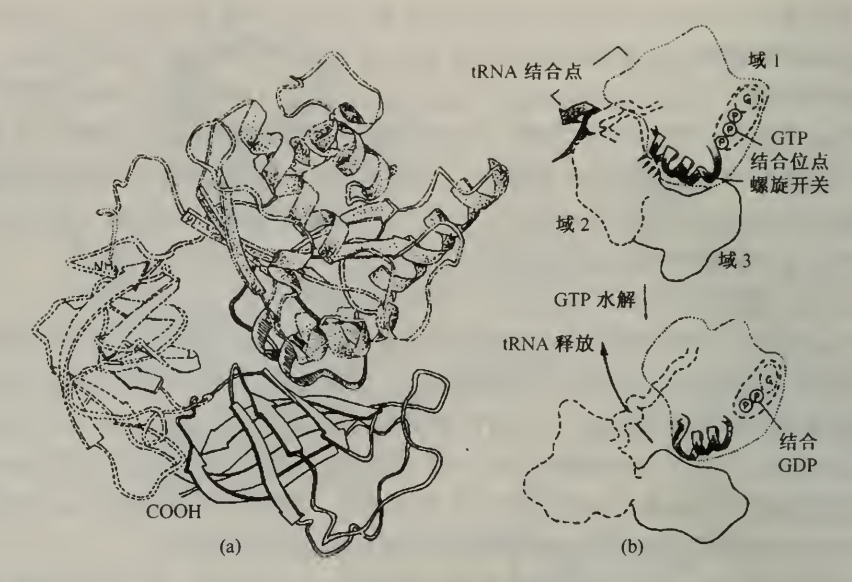


图 2-24　由 GTP 水解引起 EF-TU 的构型变化

(a) 结合了 GTP 的 EF-Tu 的三维结构，在 Ras 同源结构域上有"螺旋开关"；(b) 域 1 中的"螺旋开关"构形的改变引起域 2 和域 3 的旋转，释放出 tRNA。

蛋白质构象变化一样。例如，因为当 GTP 被水解时，大量自由能被释放，因此 EF-Tu 蛋白直接地将一个磷酸分子加到 GDP 上从而逆转 GTP 水解，这几乎是不可能的。同样的原则也可准确地用于 ATP 的水解，并且能够以一个方向长距离"步行"的蛋白质（叫马达蛋白 motor protein）通过水解 ATP 完成这样的工作。

在图 2-25 中，ATP 结合使马达蛋白从构型 1 滑动到构型 2。结合的 ATP 接着被水解，产生 ADP 和无机磷酸（Pi），引起从构型 2 到构型 3 的变化。最后释放出 ADP 和无机磷酸（Pi），驱使蛋白质从构型 3 回到构型 1。因为 1→2→3→1 的转位由水解 ATP 提供能量从而被驱动，因此在生理条件下（即 ADP 通过 1→2→3→1 循环再与 Pi 联合形成 ATP 的可能性很低），构型改变将是有效和不可逆的。因此整个循环中将以一个方向前进，在图 2-25 里，引起蛋白质向右移动。许多蛋白质以这种方式产生方向性运动，包括 DNA 解

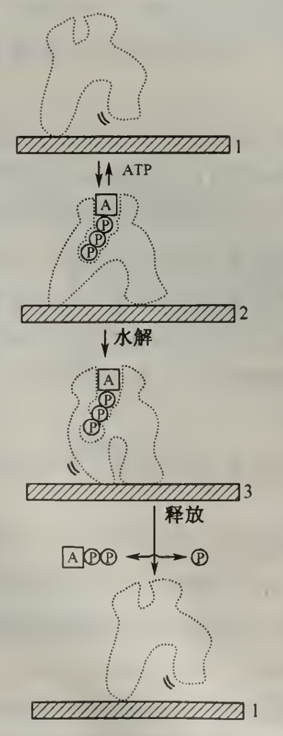


图 2-25　一个变构马达蛋白

旋酶（DNA helicase），这种酶以每秒1000个核苷酸的速度沿着DNA自我推进。

肌浆球蛋白的结构揭示了肌肉如何产生力。在ATP水解重复循环产生的能量的驱使下，通过沿蛋白纤维迅速移动的马达蛋白产生各种细胞运动（图2-25）。马达蛋白是肌浆球蛋白，它沿着肌动蛋白纤维的移动会引起细胞内的运动和肌肉的收缩。在肌浆球蛋白中也和在EF-Tu中一样，核苷酸结合位点上的一个小的移动经过放大效果的变构转移从而创造扩大的有次序的蛋白质移动。

ATP驱使的变构蛋白起离子泵作用，也能反过来合成ATP。除了产生机械力外，变构蛋白能利用ATP水解产生的能来做其他形式的工作，将特别的离子泵出或泵入细胞。例如，在所有动物细胞质膜上的Na^+-K^+-ATPase，在ATP水解驱使的构型变化的每个循环中，泵出三个Na^+，泵入两个K^+。在大多数细胞中这种ATP驱使的泵会消耗细胞所需总能量的30%。随着Na^+的不断泵出和K^+的不断泵入，保持细胞内的Na^+浓度比细胞外低得多，而细胞内的K^+浓度比细胞外的高得多，由此产生跨越细胞质膜的两种离子梯度（方向相反）。跨越细胞质膜的离子梯度能够储存能量，正像一个水坝两边的水压差能够聚集能量一样。这些能量驱使那些结合在膜上的、能做有用功的变构蛋白发生多种多样的构型变化。跨越细胞质膜的巨大的Na^+梯度可以驱使许多其他的质膜结合蛋白质泵，该泵运输葡萄糖或氨基酸进入细胞中，在Na^+涌入的同时将葡萄糖或氨基酸运送进细胞。

由ATP水解驱使的膜结合变构泵也可以从反面来工作，使用离子梯度中的能量合成ATP，事实上，跨越线粒体内膜的H^+梯度以这种方式被膜结合变构蛋白复合物——ATP合成酶使用，它合成动物细胞所需要的大部分ATP。

蛋白质中的能量-偶联变构使其行使马达（motor）、时钟（clock）、装配因子（assembly factor）或信息传感器（transducers of information）的功能。

蛋白质常常形成巨大的复合物，行使蛋白质机器的功能。在体外系统中可以看到细胞内每一个核心过程，诸如DNA复制、RNA或蛋白质的合成、空泡的出芽或者发出转送膜信号等均由10种以上的蛋白质催化。在这样的蛋白质机器中核苷三磷酸的水解（ATP或GTP）可以按次序驱使各种蛋白质的构型改变，使全部蛋白质协调移动。例如酶可以以这种方式被直接移动到它所需要执行反应的每个位置上，而不是等待反应所需的各个分散成分的随机聚集。

三、蛋白质重要的结构域

（一）结构域的化学概念

Wetlaufer于1973年根据对蛋白质结构及折叠机制的研究结果，提出了介于二级结构和三级结构之间还存在另一种结构层次即结构域（domain）。他定义结

构域为蛋白质亚基结构中明显分开的紧密球状结构区域，又称为辖区；可以通过X射线晶体衍射或电镜识别之。结构域的概念现已被生物学工作者普遍接受。的确，在一些较大的蛋白质中，往往存在两个或多个在空间上可明显区分的、相对独立的区域性结构。而对于较小的蛋白质分子，结构域和它的三级结构往往是一个意思。结构域与结构域之间常常有一段长短不等的肽链相连，并且在两个结构域之间有一明显的“颈部”或称为“凹口”，形成所谓铰链区（hinge region）。不同的蛋白质中，其结构域的数目不同。常见结构域的氨基酸残基数在100～400个之间，最小的结构域只有40～50个氨基酸残基，最大的结构域可超过400个氨基酸残基。结构域常常具有某些有趣的结构特征，比如有的结构域的特殊氨基酸含量之高异乎寻常（富含Pro结构域，酸性氨基酸结构域和富含Gly结构域等），有的结构域出现在多种蛋白质分子中（Src同源结构域SH2和SH3等），有的结构域含有特殊的二级结构组合——结构花式。

有时也可以从功能的角度给结构域下定义：结构域是蛋白质分子中具有一定活性的小区域。例如，负责催化功能的区域（激酶结构域等），或者结合功能的区域（DNA结合结构域，膜结合结构域等）。利用蛋白酶将蛋白质分子切割成最小的有活性的片段，是常用的发现功能性结构域的实验方法。另一种方法是，利用基因诱变去除蛋白质分子的部分肽段，再检测其功能。

结构域的功能性定义不如结构性定义严格。然而，如果蛋白质分子的三维结构尚不清楚，功能性结构域的发现可以提供有用的蛋白质信息；因为蛋白质活性通常依赖于一定的三维结构，而功能性结构域至少由一个（一般几个）结构性结构域组成。

（二）结构域与蛋白质结构与功能

蛋白质的生物学功能是以其化学组成和结构为基础的。一种蛋白质的生物功能的实现，不仅需要一定的化学结构，而且还需要分子的一定空间构象。下面将分化学结构与功能、蛋白质构象与功能的关系做扼要的论述。

1．蛋白质的化学结构（一级结构）与功能

蛋白质实现其生物学功能，从根本上说决定于它的化学结构。如蛋白质分子异常而导致的分子病，多肽的切割加工，蛋白质的化学修饰等都从一个侧面反映了蛋白质的化学结构与其功能的关系。

（1）一级结构的细微变化可直接影响其功能

某些蛋白质分子中个别氨基酸残基或某一肽段的被切除或置换即导致活性改变或失活。例如，胰岛素分子病是由于胰岛素分子中B链第24位的苯丙氨酸被亮氨酸取代，使胰岛素成为活性很低的分子，不能降血糖。镰状红细胞贫血是由

于血红蛋白基因中的一个核甘酸的突变导致该蛋白质分子中β链第6位的谷氨酸被缬氨酸取代。1988年，美国国立癌症研究所的Barbacid和麻省理工学院的Weinberg发现：膀胱癌细胞中的一种P21蛋白（$M_r=21\ 000$Da）与正常细胞中的这一蛋白相比，也仅仅是一个氨基酸的差异（第12位甘氨酸被缬氨酸取代），因而发生癌变。胃蛋白酶分子中的Glu-Ser-Thr肽段是表现消化力的必需结构，一经破坏，即失去活性。

(2) 多肽的切割加工

真核细胞中，负责蛋白质合成的核糖体是同内质网相连的。在核糖体上合成的分泌型蛋白质的N端都有由15～30个疏水性氨基酸组成的一段小肽，即信号肽。它的作用是使蛋白质从内质网进入高尔基体，在高尔基体内信号肽被信号肽酶切除。此时分泌出细胞的蛋白质就具有生物学活性。有些蛋白质只经过一次加工切除信号肽还不表现出生物学活性，还要经第二次加工。例如，胰岛素在含信号肽时称为前胰岛素原，切除信号肽后生成的胰岛素原还要经过加工方能成为有活性的胰岛素。

(3) 蛋白质的化学修饰

蛋白质可通过化学修饰来调控其生物学活性。化学修饰主要有磷酸化和糖基化两种形式。① 蛋白质磷酸化。蛋白激酶将ATP末端磷酸基团转移至多肽链中的氨基酸残基上。一般的蛋白激酶是将磷酸基团转移到丝氨酸和苏氨酸的残基上；癌基因编码产生的蛋白激酶则多半是酪氨酸激酶，即将磷酸基团转移到酪氨酸残基上。磷酸化可以大大提高酶蛋白的活性，如RNA聚合酶I被专一的蛋白激酶磷酸化后，其聚合活性提高了3～4倍。但是，也有酶在被磷酸化后，其酶活性会降低甚至消失。② 蛋白糖基化。许多重要的细胞表面蛋白，识别蛋白和分泌蛋白都带有一个或几个糖基。蛋白质糖基化后其构型通常会发生变化，有助于抵御蛋白酶的降解作用，并可提高蛋白质的可溶性，以及可使蛋白质准确地进入各自的细胞器。

2. 蛋白质的构象与功能

(1) 细胞内分子伴侣辅助蛋白质折叠成正确构象

在体外蛋白质一般不能正确、有效地折叠，只有少数蛋白质能在几分钟内完成折叠。在细胞内，蛋白质浓度高达100mg/ml，这种高浓度在体外将造成蛋白质沉淀；然而大约95%存在于细胞中的蛋白质显示天然构象。蛋白质分子伴侣（molecular chaperone）的广泛存在，可以解释为什么在细胞内蛋白质折叠如此有效。分子伴侣是一个新发现的蛋白质家族，广泛存在于从细菌到人的细胞中。分子伴侣能够结合各种蛋白质分子，可以参与蛋白质的一般折叠过程。分子伴侣通常以下面的两个方式帮助蛋白质分子折叠：① 结合并稳定未折叠的或部分折叠的蛋白质分子，防止它们降解；② 直接促进蛋白质分子折叠。

分子伴侣结合并稳定其靶蛋白的能力具有特异性，并且依赖于 ATP 的水解。

(2) 蛋白质的构象与功能

生物体内各种蛋白质都折叠成特定的构象是与它们各自的功能相适应的，构象改变会导致蛋白质功能改变或丧失。例如，核糖核酸酶因高浓度的尿素的影响，二硫键被破坏，肽链伸展，活性丧失。许多蛋白质行使功能时需要蛋白质内亚基与亚基之间或蛋白质与蛋白质之间相互作用形成同源或异源多聚体或大分子复合物。在相互作用时，双方或单方的结构或构象发生改变，进而影响蛋白质生物功能，如血红蛋白和原癌基因 *C-fos* 的蛋白产物等。

血红蛋白（包括人及马的血红蛋白）是由两条 α-肽链及两条 β-肽链所组成。α-和 β-肽链各分为 A、B、C、D、E、F、G、H 八段，每段的氨基酸的残基数不等。每条肽链各连接一个血红素形成一个亚基，亚基与亚基之间有 8 个盐键。由于 8 个盐键的存在，使整个血红蛋白分子的结构相当紧密，不易与氧分子结合。但当氧分子与血红蛋白中一个亚基的血红素的铁（Fe^{2+}）结合后，铁原子的半径缩小，就能进入卟啉环中的空穴中，并与它相邻的 F 肽段的组氨酸的残基（His，F8）相吸引而移动；遂使它所在的亚基构象发生大的改变，导致亚基间的盐键破裂；从而使原来结合紧密的血红蛋白分子变得松散，易于与氧分子结合。盐键的断裂也使 β-亚基的构象发生变化：排除了 E11Val 侧链对氧结合部位的空间障碍，使 β-亚基能够与氧结合；并增加与氧的亲和力，使 β-亚基与氧结合的速度加快数百倍。氧合反应最后使盐键全部断裂，整个血红蛋白分子就成为氧合血红蛋白的构象。像氧合血红蛋白这样由亚基组成的蛋白质，由于亚基构象的改变而引起其余亚基以及整个分子的构象、性质和功能的改变，就是蛋白质的变构效应。变构效应不限于血红蛋白，许多酶蛋白在催化过程中也普遍存在着这种效应。原癌基因 c-*fos* 的蛋白产物是磷酸化蛋白，有一个亮氨酸拉链区，却不能形成同源二聚体。可是，它可以与 *c-jun* 的蛋白产物中的拉链区形成异源二聚体。*c-fos* 的蛋白产物本身不能同 DNA 结合，可是一旦同 *c-jun* 的蛋白质产物形成拉链式异源二聚体后，却能与 jun-jun 同源二聚体一样具有 DNA 结合能力，且表现出更高的亲和力。

结构域作为蛋白质结构中的紧密球状结构单元，广泛存在于球状蛋白质中，其域间连接从 IgG 的松散连接到木瓜蛋白酶、溶菌酶等小分子蛋白的紧密连接。结构域作为功能单位表现在不同的结构域具有特定的功能、如结合底物、催化反应、亚基间相互作用及活性调节等。结构域作为结构和功能单位，对于功能复杂的蛋白质的构成具有重要意义。在这些蛋白质中结构域往往以组件方式相互组合从而构成具有多种不同功能的分子。多功能酶 DNA 聚合酶 I 是这种情况的典型例子（图 2-26）。

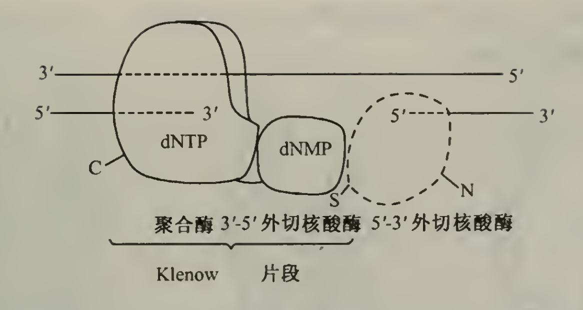


图 2-26 大肠杆菌 DNA 聚合酶的结构域结构示意图

(三) 蛋白质重要的结构域

文献中已介绍多种重要的蛋白质结构域，这里将围绕蛋白质与核酸相互作用，蛋白质与蛋白质相互作用集中介绍锌指结构，亮氨酸拉链等几种重要的蛋白质结构域，以加深读者对结构域的结构与功能的认识。

1. 蛋白质的 DNA 结合结构域

(1) 锌指结构

“锌指”（Zinc finger）结构由重复的半胱氨酸和组氨酸或重复的半胱氨酸在一个金属锌离子四周形成一个四面体的排列。由于在锌结合位点上突出的氨基酸环的形状如同手指，所以称为“指状”结构。这种蛋白质结构域是同 DNA（或 RNA）结合的部位。也有些“锌指”结构与蛋白质相互作用有关。

“锌指”结构最初是在爪蟾（*Xenopus laevis*）的 RNA 聚合酶 III 转录因子（TF IIIA）中发现的。这是九个串联重复的锌指区，每个单位大约由 30 个氨基酸残基组成。许多种调控蛋白质都有数目不等的锌指结构。含有锌指的蛋白与 DNA 的结合方式可能互不相同，有些情况下锌指结构中的氨基酸残基参加对 DNA 顺序的识别；有些蛋白的锌指却只是促进与 DNA 的非特异性结合，而顺序的识别是由蛋白质分子中的其他部分进行的。

锌指有两种类型：C_2H_2，C_4；其中，C 代表 Cys，H 代表 His。

① C_2H_2 锌指。C_2H_2 锌指是最普遍的真核转录因子 DNA 结合结构域之一，迄今发现的含有 C_2H_2 锌指结构的高等真核生物蛋白质已超过 1000 种。典型的 C_2H_2 约由 23 个氨基酸残基组成。其中，保守残基是结合锌离子的 2 个 Cys 和 2 个 His，以及 1 个 Phe 和 1 个 Leu。锌离子位于 2 个 Cys 和 2 个 His 形成的四面体结构的中心（图 2-27）。C_2H_2 锌指一般为连串重复排列，锌指之间通常有 7～8 个氨基酸残基。爪蟾的 5S RNA 转录因子 TF IIIA 有 9 个连续的锌指；果蝇调节因子 ADRI 的 DNA 结合结构域很少，只有 2 个 C_2H_2 锌指；尽管 C_2H_2 锌指蛋白

有多个锌指，但不一定每个锌指都接触 DNA。例如，人 GL1 蛋白有 5 个锌指，第一个锌指就不接触 DNA。

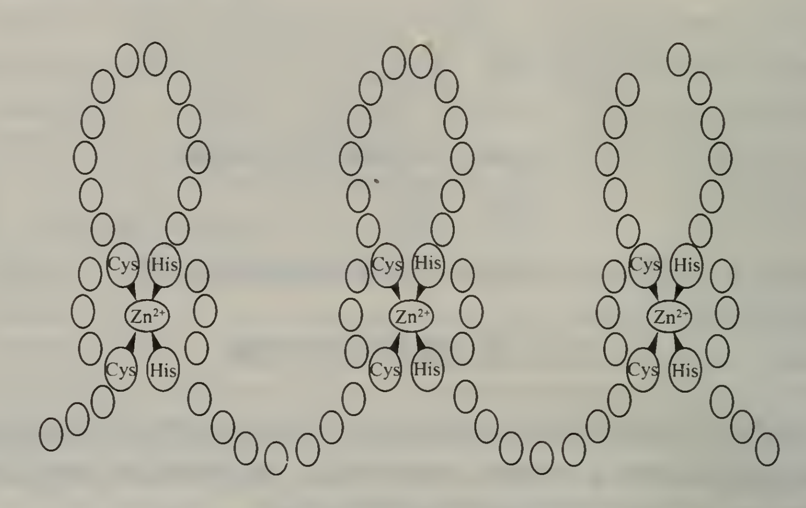


图 2-27　C_2H_2 型“锌指”蛋白结构

C_2H_2 锌指-DNA 晶体结构显示，每个锌指的 N 端部分形成一对反向 β-折叠；而 C 端部分形成一个 α 螺旋同 DNA 的大沟结合，与大约 5 个核苷酸相互作用，从而起到识别并结合 DNA 序列的功能（图 2-28）。

② C_4 型锌指。C_4 型锌指蛋白已发现 100 多种，主要是类固醇受体家族成员组成的转录因子。C_4 型由数目不等的 Cys 同金属锌离子螯合。C_4 类的结构同 C_2H_2 相似：$C\text{-}X_2\text{-}C\text{-}X_{13}\text{-}C\text{-}X_2\text{-}C$，由 13 个氨基酸残基构成一个环。$C_5$ 的结构是 $C\text{-}X_5\text{-}C\text{-}X_9\text{-}C\text{-}X_2\text{-}C\text{-}X_4\text{-}C$，$C_6$ 的结构则是 $C\text{-}X_2\text{-}C\text{-}X_6\text{-}C\text{-}X_9\text{-}C\text{-}X_2\text{-}C\text{-}X_6\text{-}C$。$C_5$ 和 C_6 可能有多个金属离子与 Cys 螯合。C_4 锌指蛋白通常仅含有少量锌指，识别 DNA 短回文序列或同向重复序列，以同源或异源二聚体的形式行使功能，而 C_2H_2 锌指蛋白含有多个重复排列的锌指，并且以单体形式结合 DNA。

糖皮质激素受体和雌激素受体都有 2 个 C_4 锌指，彼此间隔 7 个氨基酸残基，锌离子位于每个锌指的 Cys 四面体结构中心。这两个锌指的两个 α 螺旋折叠成球形，与之接触的是 β 折叠。第一个锌指右侧的氨基酸残基（α 螺旋）结合于 DNA 大沟，而第二个锌指左侧的氨基酸残基调节其形成二聚体的能力。在糖皮质激素受体锌指第一个锌指的右侧有 2 个特异的氨基酸残基 Gly 和 Ser，而雌激素受体的相应位点是 Glu 和 Gly。这两个残基以及附近间隔区的一个氨基酸残基决定受体 DNA 结合位点的特异性，使它们分别结合于特异的应答元件（GRE 和 ERE）。这两种受体基因碱基序列的区别主要集中在第一个锌指，特别是上述 3 个氨基酸残基。如果进行“特异性交换”（specificity swap）实验，使雌激素受体的第一个

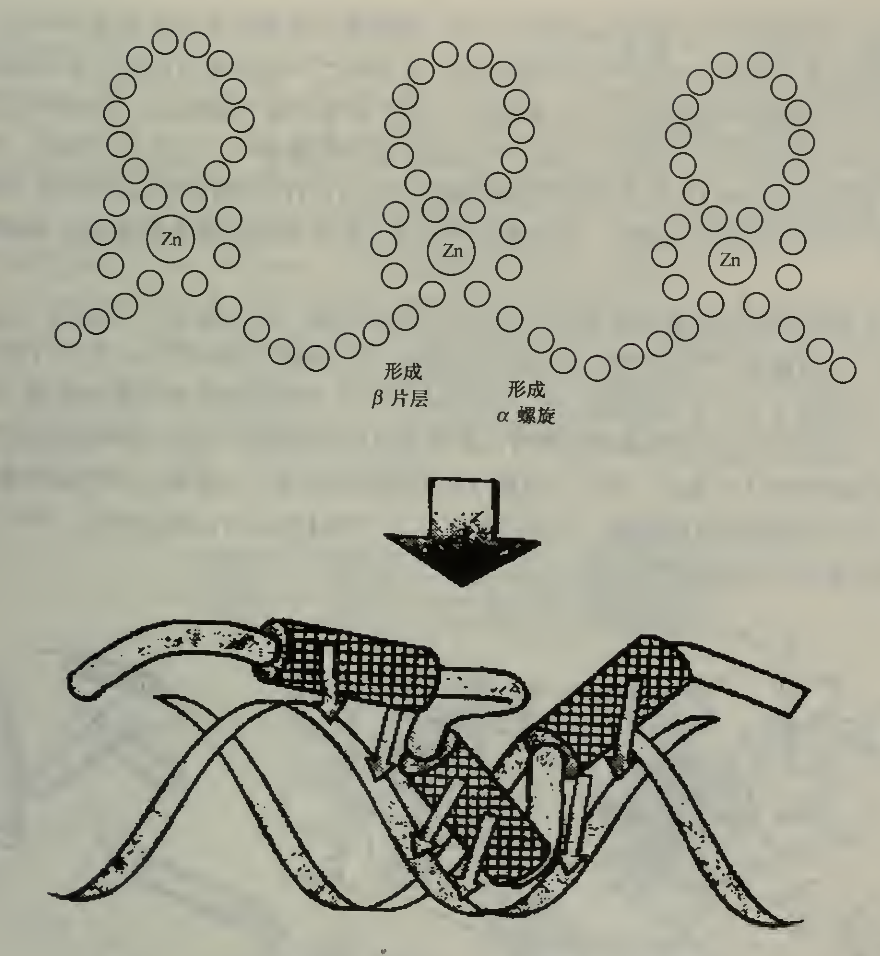


图 2-28 “锌指”可形成 α 螺旋插入 DNA 大沟，另一侧呈 β 片层结构

锌指缺失，代之以糖皮质激素受体的第一个锌指；结果显示新蛋白识别糖皮质激素应答元件（GRE），这证明两种受体的第一个锌指决定 DNA 识别特异性。

C_6 锌指蛋白以酵母转录因子 Gal4 为代表。Gal4 以二聚体形式结合于 DNA，两个单体通过它们的 α 螺旋一侧的疏水侧链相互作用，形成卷曲线团（coiled coil）。类似的相互作用也出现于亮氨酸拉链蛋白二聚体。

一个蛋白质分子也可能存在两种指状结构。例如，类固醇受体就含有两个显然是相互无关的指状结构，分别由两个外显子编码。

（2）螺旋-拐角-螺旋

这是第一个经过详细研究的 DNA 结合结构域，它是许多原核生物调节蛋白中蛋白质与 DNA 相互作用的重要结构，类似的结构域也存在于真核调节蛋白中。螺旋-拐角-螺旋（helix-turn-helix，HTH）有两个 7～9 个氨基酸残基组成的

α螺旋，中间由一个含脯氨酸的拐角分开。该拐角不是随机的构象而是一种特定的结构拐角（turn），由4个氨基酸组成，其中第二个通常是甘氨酸。这个拐角与第一个α螺旋一起将第二个α螺旋以一定方向置于蛋白质表面，使其可进入DNA分子大沟内部。因此，这第二个α螺旋是识别螺旋（recognition helix），它通常含有许多能和DNA相互作用的氨基酸。这种HTH结构通常具有约20个氨基酸，并且本身不是很稳定，但它是较大的DNA结合结构域中起关键作用的部分。

λ噬菌体的阻遏蛋白就有这种DNA结合结构域。在阻遏蛋白N端有5段α螺旋，每段螺旋之间折转成一定角度相连接，其中两段负责同DNA结合（图2-29）。α螺旋3由9个氨基酸组成，与前面的由7个氨基酸组成的α螺旋形成一个角度。α螺旋3通过氨基酸侧链和DNA碱基之间的氢键同DNA序列相结合，3个氨基酸识别3个碱基，所以α螺旋3被称为识别螺旋。α螺旋2则是通过氢键同DNA的磷酸骨架相接触。这种相互作用对于同DNA结合是必需的，但并不控制对靶序列识别的专一性。

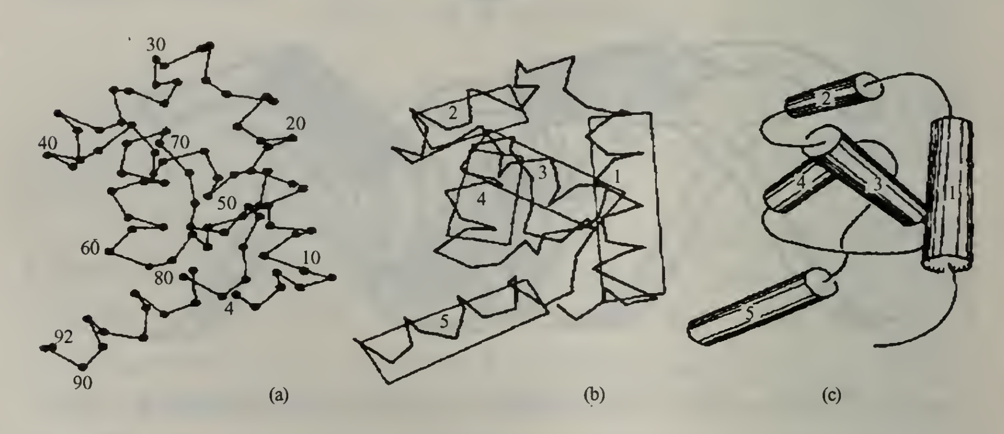


图2-29　用X射线衍射法测定的λ阻遏蛋白N端结构域的结构

（a）示蛋白链的折叠（二维），每一个点代表一个氨基酸的α-碳的位置，链上有5个α螺旋区；（b）示用矩形框出的5个α螺旋区；（c）结构域的示意模型，圆柱体代表α螺旋区，线代表非螺旋区，α螺旋3从表面向外突出。

（3）同质异形结构域

同质异形结构域（homeodomain，HD）也叫同源框结构域或同源域，它由60个氨基酸残基组成，而且这个肽段里的氨基酸组成在进化上是高度保守的，从果蝇、小鼠到人很少发生改变。编码这种结构域的DNA顺序叫做同质异型盒或同源异形框（homeobox）。同质异型盒相关序列和HD蛋白，广泛分布在从酵母到哺乳动物的所有真核生物，仅果蝇就有60多种HD蛋白。HD蛋白对真核生物发育和分化调控具有十分重要的作用。

不同基因的同质异型盒序列的保守性可以改变。果蝇 3 号染色体有两个主要的同质异形基因簇 ANT-C 和 BX-C：ANT-C 有 5 个基因，负责头部-前脑的发育；BX-C 有 3 个基因，负责后脑-腹部的发育，这些基因的同源异形框相当保守，其序列有 80%～90%的碱基对相同。但在其他含同源异形框的果蝇基因中，同源异形框之间的相似程度很低。动物有几种转录因子含 HD，但果蝇和哺乳动物 HD 之间的氨基酸顺序保守性明显下降［图 2-30，(a)］。虽然酵母和果蝇 HD 之间的相似性只有大约 30%，但其识别 DNA 的方式却相同。

HD 识别 DNA 序列的机制与 λ 阻遏蛋白的 HTH 相似。HD 的 60 个氨基酸残基组成 3 个 α 螺旋，其中最保守的 α 螺旋 3 结合于 DNA 双螺旋的大沟［图 2-30，(b)］；而 α 螺旋 1 和 α 螺旋 2 反向平行排列，位于双螺旋外面。在结构上，HD 和原核阻遏蛋白 HTH 的区别在于：HD 的 α 螺旋 3 长度为 17 个氨基酸残基，而 λ 阻遏蛋白的 α 螺旋 3 只有 7 个氨基酸。

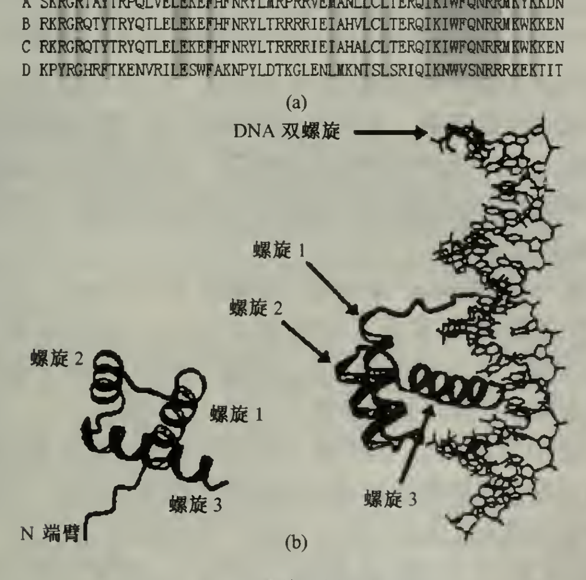


图 2-30　同源框结构域（HD）

(a) HD 模体经常出现于发育调节因子分子，例如，A 为哺乳动物小鼠 *MO-10* 基因、B 为两栖动物蛙 *MM3* 基因、C 为昆虫果蝇 *Antp* 基因和 D 为酵母 *MATα2* 基因，其氨基酸顺序相似，尤其是 α 螺旋 3 最保守，但哺乳动物和果蝇 HD 之间的相似程度明显下降；(b) HD 模体由 3 个 α 螺旋组成。α 螺旋 3 插入 DNA 大沟，α 螺旋 1 和 α 螺旋 2 反平行排列，HD 的 N 端臂与 DNA 的小沟接触。

2．蛋白质的 RNA 结合结构域

RNA 结合蛋白的主要 RNA 结合结构域有：

核糖核蛋白结构域［ribonuceloprotein（RNP）domain］包括4个β链和2个α螺旋，其顺序是β-α-β-β-α-β。两个中央β链构成与RNA分子的主要接触。RNP结构域是最普遍的RNA结合结构，现已在250种以上的蛋白质中发现。由于此类RNP只存在于细胞核内，因此可能与RNA在核内的滞留有关。此外，由于这种RNP与被剪切的内含子3′端的多聚嘧啶区及其下游polyA拼接点的RNA序列有关，所以可能参与剪接和polyA拼接的调控。

RGG盒。RGG盒是由20～25个氨基酸组成的RNA结合单元，常与其他类型的RNA结合结构域共同存在，最早发现于hnRNPU的RNA结合结构域中。这一序列的特征是较密集的、多拷贝的Arg-Gly-Gly（RGG）重复序列，间隔以其他氨基酸，最常见的是芳香族的氨基酸。不同RNA结合蛋白中RGG重复序列可达6～18次。在核仁蛋白中，4个RNP基序可特异地与核糖体RNA前体结合，而核仁蛋白中唯一的RGG盒区可将亲和力提高10倍。

双链RNA结合结构域（double-strande RNA binding domain，dsRBD）与RNP结构域相似，但其结构是α-β-β-β-α。它的RNA结合作用存在于结构两端的α和β间，此结构发现于结合双链RNA的蛋白中，而且发现愈来愈多的RNA结合蛋白含有一个或多个拷贝的双链RNA结合区。它大约由70个氨基酸残基组成，含有碱性和疏水性氨基酸，散布在DSRM区内。哺乳动物细胞对病毒感染的重要应答产物是一种干扰素诱导的蛋白激酶，它含有两个DSRM，能与RNA双链结合。当被病毒dsRNA激活时可封闭宿主蛋白的合成，因此又名为dsRNA激活抑制物（DAI）。

k-同源结构域（k-homology domain）结构是β-α-α-β-α-α，结合功能存在于α螺旋对中。这种结构并不常见，但存在于至少一个结合蛋白中。在人的hnRNP-K蛋白中发现，已知功能的KH结构蛋白都以与RNA结合的形式存在，而且最近有证据表明KH基序是直接接合RNA所必需的。

在某些蛋白质中结合DNA的同源异形结构域也可能有RNA结合活性。有一种核糖体蛋白用相似于同源异形结构域的结构结合RNA；一些同源异形结构域蛋白，如黑腹果蝇的Bicoid既可结合DNA，又可结合RNA。

锌指也有RNA结合结构域的功能。例如，一些蛋白通过锌指结合真核mRNA，起到翻译阻遏子的作用。TFIII-A为含有9个锌指结构的蛋白质，能与5S rRNA及其基因结合，其功能与5S rRNA运输或RNA聚合酶III转录有关。

3．蛋白质-蛋白质相互作用结构域

（1）螺旋-环-螺旋

螺旋-环-螺旋（helix-loop-helix，HLH）结构域含有40～50个氨基酸残基，在2个两性α螺旋之间有一个长度可变的连接区（loop），每个α螺旋区长15～16个氨基酸，其中有几个氨基酸残基是保守的。通过α螺旋之间相互作用，

HLH 蛋白能够形成同源或异源二聚体（图 2-31），环的存在可以保证两个 α 螺旋独立地起作用。

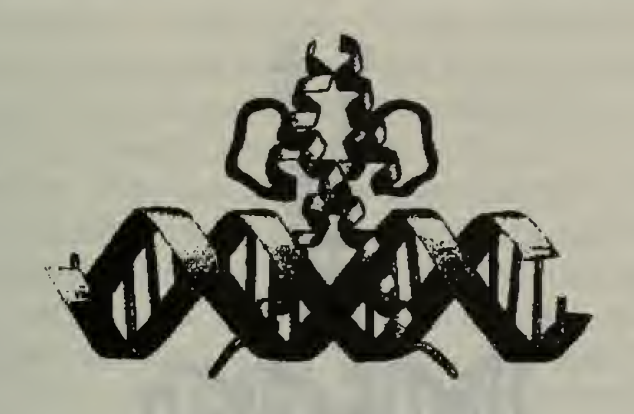


图 2-31　HLH 蛋白二聚体结合于 DNA
两个 HLH 蛋白单体的两性 α 螺旋之间相互作用，形成 HLH 蛋白二聚体，在每个单体中，两个两性 α 螺旋之间的肽段呈柔性的环状结构。

在 HLH 结构域附近，绝大多数 HLH 蛋白富含大量碱性氨基酸，在 15 个氨基酸残基中约有 6 个是保守的，这个区域是结合 DNA 所必需的。符合这一特点的 HLH 蛋白叫做“bHLH 蛋白”，包括 E12 和 E47 蛋白、原癌基因 *c-myc* 的表达产物 Myc 蛋白等。有的 bHLH 蛋白的存在没有细胞特异性，如哺乳动物的 E12/E47 和果蝇的 Da 等。但有的 bHLH 蛋白具有组织特异性，如哺乳动物的 MyoD 和果蝇的 Ac-s 等。

bHLH 蛋白二聚体的形成及其结合 DNA 的难易和二聚体的伴侣有关。例如 E47 同源二聚体和 E12-E47 异源二聚体都容易形成，并且具有较强的结合 DNA 能力。而 E12 同源二聚体虽然容易形成，但难于结合 DNA。MyoD 同源二聚体正好相反，难于形成却易于结合 DNA。

另一种类型的 HLH 蛋白没有碱性区，并且含有破坏其 DNA 结合功能的氨基酸残基 Pro，如 ID 蛋白。这种类型的 HLH 蛋白，具有像 bHLH 蛋白一样的形成二聚体的能力，但是二聚体中有一个亚基没有碱性区，这个亚基不能特异地结合 DNA 。

（2）亮氨酸拉链

亮氨酸拉链（leucine zipper）是一种重要的蛋白质-蛋白质相互作用的结构域。其每 7 个氨基酸中的第 7 个氨基酸是亮氨酸，亮氨酸是疏水性氨基酸，排列在 α 螺旋的一侧，所有带电荷的氨基酸残基排在另一侧。当 2 个蛋白质分子平行排列时，亮氨酸之间相互作用形成二聚体，即形成拉链（图 2-32）。在拉链式的蛋白质分子中，C 端的亮氨酸拉链本身并不能结合 DNA，结合 DNA 的是 N 端的

碱性氨基酸区。因此，这种结构域又叫碱性拉链（basic zipper），即 bzip。亮氨酸拉链广泛存在于真核生物调控蛋白中，已发现几个原核蛋白也含亮氨酸拉链。



图 2-32 酵母转录因子 Gcn 4 的亮氨酸拉链
两个富含亮氨酸的 α 螺旋像拉链一样相互作用，形成的 Gcn 4 二聚体呈剪刀状夹住 DNA，另一个与大沟接触的 α 螺旋富含碱性氨基酸残基。

亮氨酸拉链（bzip）蛋白可以形成同源或异源二聚体。不同的 bzip 蛋白所含有的 Leu 重复单位数目不等。C/EBP 二聚体结合于 CAAT 盒或 SV40 增强子，它的亮氨酸拉链有 4 个 Leu 重复单位。转录因子 API 由 Jun-Fos 异源二聚体组

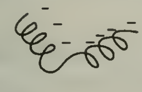
成，最初发现它结合于 SV40 增强子，它的亮氨酸拉链有 5 个 Leu 重复单位。

(3) SH2 和 SH3 结构域

活化 RTK 的靶蛋白（细胞内信号蛋白）的结构与功能形形色色，但是它们一般都含有两个高度保守的非催化性结构域 SH2 和 SH3。SH2 结构域能识别并结合活化 RTK 的 p-Tyr，以及其他含有 p-Tyr 的细胞内信号蛋白分子。不同 RTK 靶蛋白的 SH2 结构域的三维结构十分相似，它们结合于 RTK 的 p-Tyr 周围特异的氨基酸顺序，而每个 SH2 独特的氨基酸顺序决定它结合于哪些特异的 p-Tyr。例如，Src 蛋白紧紧的结合于含有核心顺序 p-Tyr-Glu-Glu-Ile 的任何多肽，这四个残基形成的空间结构正好和 Src 蛋白的 SH2 多肽结合位点相匹配，它们的结合类似于一个两相的电源插头（即核心顺序中的 p-Tyr 和 Ile 侧链）和电源插座（SH2 结构域）。SH2 肽结合位点有一个能容纳 p-Tyr 侧链的孔穴，其中经常有 2 个 Arg 和 1 个 Lys 残基，它们和 p-Tyr 之间产生较强的静电引力。此外，SH2 还有一个容纳 Ile 侧链的疏水性孔穴。不同的 SH2 结构域的结合特异性，主要取决于组成其疏水性孔穴的氨基酸残基，这些氨基酸使之结合于和 p-Tyr 相邻的特异氨基酸顺序上。

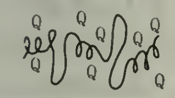
SH3 结构域的蛋白结合位点大约有 10 个氨基酸残基，其中含有保守序列 Pro-X-X-Pro。人们对 SH3 结构域功能的认识，来自对一种小接头蛋白的研究所获得的启示。人寄生虫线虫 *C. elegans* 有一种小接头蛋白 Sem-5，如果它的基因发生突变，将抑制几种 RTK 信号传递途径，对线虫发育造成很大的影响。有趣的是，Sem-5 蛋白的 SH2 结构域编码序列突变产生的抑制效应，相当于它的两个 SH3 结构域中的任何一个发生突变，这就提示：这两个 SH3 结构域是结合下游信号蛋白所必需的。事实上，在绝大多数的动物细胞中都有和 Sem-5 蛋白同源的接头蛋白；有证据表明，它们将活化的 RTK 和重要的下游信号蛋白 Ras 连接起来。

GAL4 酸性结构域



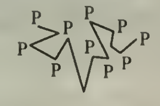
DSAAAHHDNSTIPLDFMPRDALH
GFDWSEEDDMSDGLPFLKTDPNNNGF

SPI 谷氨酰胺结构域



QGQTPQRVSGLQGSDALNIQQNQTSGGSLQAGQQKEGEQNQQTQQQQL
QPQLVQGGQALQALQAAPLSGQTFTTQAISQETLQNLQAVPNSGPIIIRT
PTVGPNGQVSWQTLQLQNLQVQNPQAQTITLAPMQGVSLGQ

富含脯氨酸结构域



PPHLNPQDPLKDLVSLACDPASQQPGRLNGSGQLKMPSHCLSAQMLPPPP
GLPRLPPATKPATTSEGGATSPSYSPPDTSP

图 2-33　几种蛋白质转录活化结构域

介导蛋白质间相互作用的结构域有多种，除上述三类结构域外，研究较清楚的还有：富谷氨酰胺结构域（图 2-33），富脯氨酸结构域（图 2-33），酸性结构域（图 2-33）等。从名字可以反映出这些结构域中氨基酸残基的组成特点。

四、生物大分子相互作用和复杂聚集物的结构

在前面章节里我们讨论了蛋白质及核酸的结构，也介绍了具有两个或更多的相同或不同亚基的蛋白质。核酸中也是如此，尽管其亚基的数目很小。例如双链 DNA 分子有两个亚基，而反转病毒 RNA 的亚基数比其他任何核酸的亚基数都多，有四个亚基。

有机体的细胞内外普遍存在不同生物大分子之间的相互作用。例如，核酸常和蛋白质结合在一起形成 DNA-蛋白质复合物（染色体），细胞外的病毒核酸被包装在蛋白质外壳内，骨胳和软骨是蛋白质和其他生物大分子复杂的装配体。蛋白质也能和脂相互作用产生膜，如将细胞内含物与外环境隔开的膜以及将细胞内成分相互隔开的膜。多糖形成像细菌和植物细胞壁那样极复杂的结构。

对这些复杂结构及其形成的研究叫结构生物学。有少数结构几乎完全了解清楚了，而许多其他结构正在积极地研究中。本节只讨论少数几种生物大分子聚集物的结构以阐明生物大分子相互作用的重要性及作用的基本原理。

1. 一种多蛋白装配体——胶原蛋白

胶原蛋白是哺乳动物中最丰富的蛋白质，也可能是世界上最普通的蛋白质。它是腱、软骨、骨骼、皮肤和血管的主要蛋白质成分。其主要功能是提供具强大抗张强度的纤维。胶原蛋白是一个由亚基联合形成的三聚蛋白装配成的蛋白多聚体，有高度的组织（organization）；三聚蛋白进一步联合形成纤维。胶原蛋白由成纤维细胞合成，以称为原胶原（procollagen）的前体被分泌到胞外，然后按以下步骤装配：

原胶原细丝（procollagen filament）　　(0～4nm)
↓
原小纤维（protofibril）　　(11～15nm)
↓
微小纤维（microfibril）　　(30～200nm)
↓
胶原蛋白小纤维（collagen fibril）　　(0.2～0.4μm)
↓
纤维（fiber）　　(1～10μm)
↓
腱、骨骼、软骨等（肉眼可见）

(1) 胶原蛋白的氨基酸组成和顺序

装配成胶原蛋白的多肽链的氨基酸组成及顺序有几个显著特征：

①每条链含有近 1000 个氨基酸，1/3 的氨基酸是甘氨酸；这个值比普通蛋白质高很多，普通蛋白质中甘氨酸只占 5%。胶原蛋白多肽链的氨基酸顺序可表示如下：

…Gly-A-B-Gly-C-D，Gly-E-F-Gly-G-H…

在上述表示中，A、B、C…是甘氨酸外的其他氨基酸。

②1/4 的氨基酸是脯氨酸或脯氨酸的衍生物羟脯氨酸。脯氨酸在其他蛋白质中很少遇见，而羟脯氨酸没有在其他蛋白质中发现。

③许多赖氨酸羟基化。

④甘氨酸-脯氨酸-羟脯氨酸顺序反复出现，在绝大多数其他蛋白质中是很罕见的，这种氨基酸组成及其排列方式与胶原蛋白重要的性质相对应。

(2) 一种三链螺旋单位——原胶原蛋白（tropocollagen）

除了胶原蛋白合成的某一阶段外，胶原蛋白的单链单体在天然条件下是不存在的，只存在由单体联合形成的原胶原蛋白。

各个多肽链中存在大量脯氨酸，从而影响了原胶原蛋白分子结构。脯氨酸是一种不能形成典型肽键的氨基酸，这些键的结合使柔韧降低因而使肽链更趋于伸展。由于链的伸展，肽链中的许多可相互作用的基团暴露在外，使得各个链间能通过链内的氢键相互作用产生三链螺旋，这是形成具有强大抗张力纤维的第一步。氢键和范德华力将各个链拉在一起形成一个非常紧密的复合物，在此复合物中，脯氨酸和几乎所有的其他侧链氨基酸向外而甘氨酸向内。稳定原胶原三螺旋的力除了范德华力、氢键和链间的共价交联外，三螺旋的结构还要求三联体的每第三个位置必须是 Gly。三螺旋内侧很拥挤，没有容纳庞大侧链氨基酸的空间；而甘氨酸没有侧链（仅是 H 原子），因此甘氨酸能使其结构最大限度地紧密并且增加其强度。

(3) 原胶原单位相互作用形成纤维

原胶原蛋白分子在胶原纤维中都是有规则地按 1/4 错位、首尾相接、并行排列组成纤维束。由于原胶原肽链上残基所带电荷不同，因而电子密度不同，通过 1/4 错位排列便形成间隔一定的电子密度区域，而呈现横纹区带（图 2-34）。原胶原蛋白分子如果不是按 1/4 错位排列，而是每对分子的头及尾对齐排列，那么外力很易将该纤维打断；而 1/4 错位排列使得纤维不易被打断。

此外，胶原蛋白特殊的氨基酸——羟脯氨酸和羟赖氨酸在原胶原蛋白相邻的多肽链之间和相邻的原胶原蛋白链之间形成共价交联（图 2-35），而加固其 1/4 错位排列。很显然，胶原蛋白单体的许多结构特征联合在一起形成一个具有巨大抗张力强度的大分子装配体。

胶原蛋白的结构特征担负另一个特别功能。在每个以 1/4 错位排列连续的原

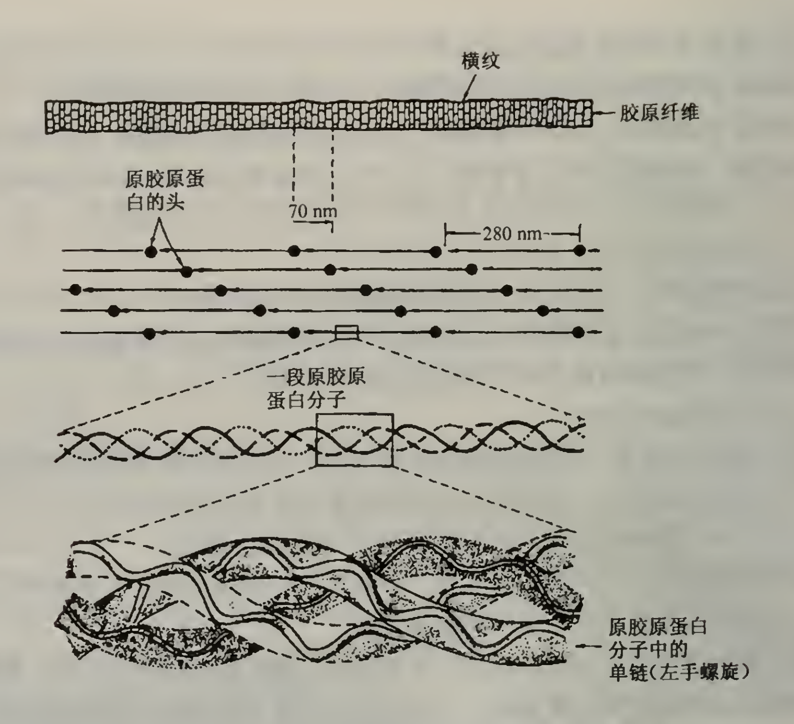


图 2-34　胶原纤维中原胶原蛋白分子的排列

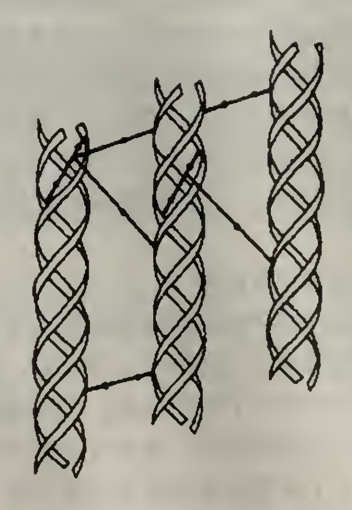


图 2-35　胶原蛋白分子链内及链间的交联

胶原蛋白分子的头尾之间是一个空隙，这种空隙在骨骼形成中起作用。骨骼是由羟基磷碳石结晶在胶原纤维上沉积而形成的，结晶过程可以在体外进行。发现其在胶原纤维上以 670Å 的间隔起始形成微结晶，这个距离正好是一个胶原蛋白空

隙的大小。结晶不在人工合成的非 1/4 错位排列的胶原纤维的空隙处形成。

2. 复杂的 DNA 结构

主要简单介绍大肠杆菌染色体和真核细胞染色体及染色质的结构。大肠杆菌和大多数细菌的基因组只含一个环状超螺旋分子，环状分子的总长度大约 1300μm。一个圆棒状的细菌的直径和长分别是 1μm 和 3μm，因此很显然细菌 DNA 在一个细胞里是高度折叠的。细菌 DNA 分子和蛋白质、RNA 相互作用形成紧密的结构即细菌染色体。这种紧密结构有两个特点：①在外围全部是 DNA 超螺旋环；②在近染色体中心位置是密度区，称作骨架（scaffold）（图 2-36）。细菌 DNA 采取多个超螺旋折叠方式将巨大的 DNA 分子压缩在一个小的细菌细胞里。当 DNA 分子上产生一个切口，将解开一个折叠中的超螺旋；产生两个切口，将产生两个松弛环（图 2-36）。DNA 旋转酶（DNA gyrase）在 DNA 复制中起重要作用，负责形成超螺旋。

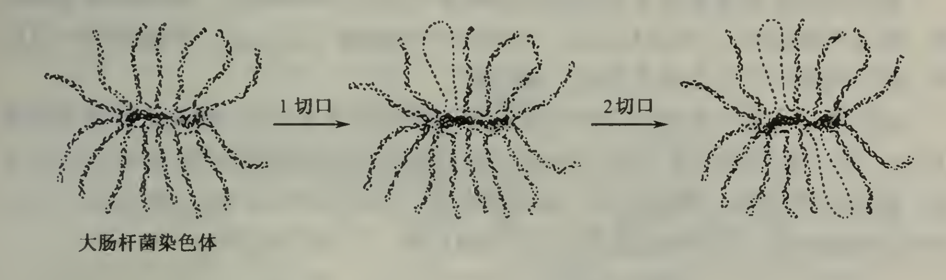


图 2-36　大肠杆菌染色体高度折叠超螺旋

真核细胞中的核内 DNA 均形成一定数目的染色体，每个染色体含有单个巨大的 DNA 分子。例如果蝇的一个染色单体含一条分子质量大于 10^{10} U、长达 1.2cm 的 DNA 分子。染色体的结构比 DNA 分子的结构紧密得多，DNA 分子不是自我折叠成像染色体那样致密的结构。真核染色体中的 DNA 分子与组蛋白结合，形成的复合物叫染色质（chromatin）。主要有五类组蛋白：H1，H2A，H2B，H3 和 H4。这五种蛋白质的性质见表 2-4。

细胞中染色质的结构随着细胞生长周期而改变。静止期的细胞中的染色质是松散的，并且充满整个细胞核。当 DNA 刚开始复制，染色质浓缩大约 100 倍，同时染色体形成。染色质是细胞间期核中 DNA、组蛋白、非组蛋白以及少量 RNA 所组成的一串念珠状的复合体。染色体具有四级结构，其中核小体是其一级结构。每个核小体包括有 200bp 的 DNA 链和 8 个组蛋白分子（包括 4 种主要组蛋白 H2A、H2B、H3 和 H4，每种两个）。这些球状组蛋白分子互相挤在一起，呈小圆盘状，DNA 链就绕在小圆盘之外，缠绕 1.75 圈，由 146bp 组成；小圆盘之间由组蛋白 H1 连接。

表 2-4 组蛋白类型

类型	Lys/Arg 比	氨基酸数	分子质量/kDa	位置
H1	20.0	215	21.0	连结处
H2A	1.25	129	14.5	核心
H2B	2.5	125	13.8	核心
H3	0.75	135	15.3	核心
H4	0.79	102	11.3	核心

核小体链呈螺旋形缠绕，形成超微螺旋，称为“螺线管”（solenoid）。这种超微螺旋的直径约为 30nm，内径 10nm，螺距为 11nm，为中空呈管状结构。这种螺旋管的每一转由 6 个核小体组成，它是染色体的二级结构。

螺线管的纤丝沿着它中央的蛋白质轴发射出大小不等的环，像灯刷染色体那样。经过观察证明，染色体是由一系列的环状的域（domain）即超螺线环组成的。这种结构可以说是染色体的三级结构。

染色体（实际上指染色单体）的形成一般认为是由 DNA 盘绕成紧密的螺线（helix）和一系列螺线管纤丝（solenoidal fiber）的环状域密集地组装成染色单体，就是四级结构。通过这样一系列的过程，染色体中的 DNA 被压缩了大约 6000 到 8000 倍。下面表述左为 DNA 压缩比率，右为压缩方法。

DNA 双螺旋
1/7　　螺旋化（H2A，H2B，H3，H4）
核小体
1/6　　螺旋化（H1）
螺线管
1/30　　折叠和螺旋化
超螺线环
1/5　　折叠和螺旋化
———
合计 1/6300　　染色单体

如将染色体中的 DNA 和组蛋白除去就是留下的蛋白质所形成的骨架。

3. DNA 与一个识别专一碱基序列的蛋白质的相互作用

我们知道在染色质的形成中，组蛋白只识别普通的 DNA 结构而不是与专一的碱基序列结合；然而许多蛋白质只对特定的碱基序列结合，本书中我们将遇到很多这类蛋白质。这些蛋白质有三种类型：①聚合酶，它们在特定的碱基位点起始 DNA 和 RNA 的合成；②调节蛋白，打开或关闭特定基因的活性；③限制性

内切核酸酶，它们能识别 4 或 6 个核苷酸序列，并在两个相邻的核苷酸之间打开磷酸二酯键。

在许多 DNA-蛋白质复合物中，蛋白质结合的专一序列是知道的。这里特别讨论一个例子，以便使我们了解被识别的 DNA 序列的特殊性质和结合的模型。

现以 Cro 蛋白对 λ 噬菌体 DNA 的结合为例。大肠杆菌噬菌体产生的 Cro 蛋白在 λ 噬菌体的生活周期中是一个重要的调节蛋白，其重要性就是因为它对 DNA 分子上的特殊碱基序列进行结合。

Cro 是个小分子质量蛋白质，由 66 个氨基酸组成，它识别并结合 λDNA 上 17 个碱基长的专一序列。此蛋白质的氨基酸序列和它识别的 17 个碱基序列都已知，这些资料使得蛋白质与 DNA 相互作用是通过结构特征这个假设得到证实。

Cro 蛋白的三维结构已经用 X 射线衍射技术测定，如图 2-37。与大分子蛋白相比，它是一个简单的结构。含有三个反向平行的 β 层（2～6、39～45 和 48～55 残基）和三个 α 螺旋（7～14、15～23 和 27～36 残基），由于 63～66 残基无次序，没有在图 2-37 中表明（它们与 Cro-DNA 相互作用没有关系）。

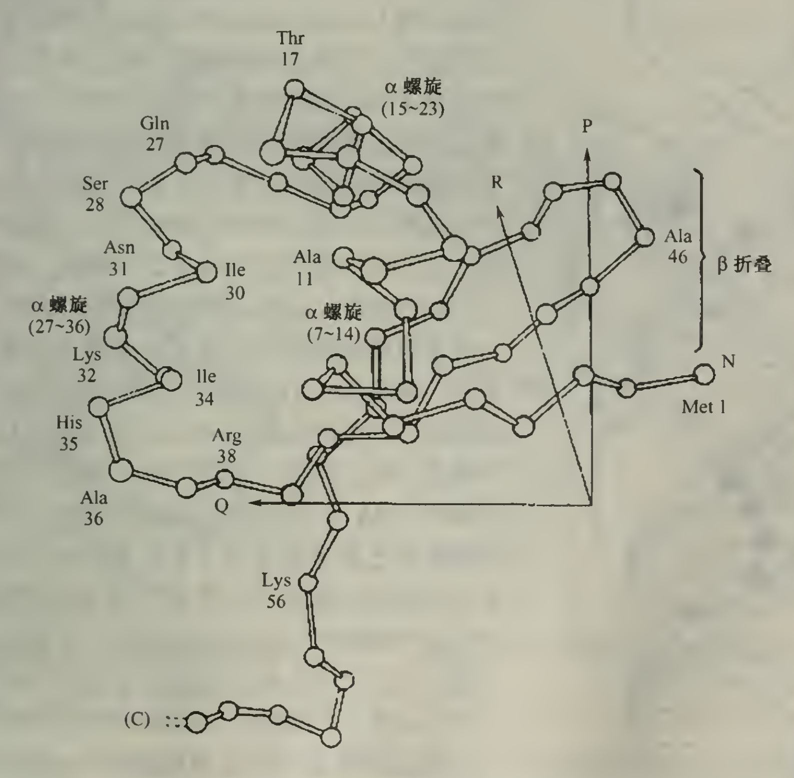


图 2-37　λ Cro 蛋白的三维结构

λCro 由 α 螺旋和 β 折叠组成，箭头指的 P、Q 和 R 代表其他三个侧面的轴。

图 2-38 表示出 λ DNA 上与 Cro 蛋白结合的 17 个碱基的顺序。

T A T C A C C G C A A G G G A T A
A T A G T G G C G T T C C C T A T

图 2-38　λ Cro 蛋白的结合位点之一的碱基序列

通过 3 种技术研究 Cro 蛋白质与 DNA 之间接触处的某些特点。第一个技术是纯化 Cro 蛋白和 DNA，并且在能引起 Cro-DNA 形成复合物的条件下将其混合；这些 Cro-DNA 复合物再用硫酸二甲酯处理，硫酸二甲酯在鸟嘌呤环的 N-7 位点和腺嘌呤环的 N-3 位点加上甲基。然后使 Cro-DNA 解离，除去 Cro 蛋白后分析这段 DNA 的顺序，确定 G 和 A 的位置。那些未甲基化的 G 和 A 就在 Cro-DNA 复合物中，因为 Cro 蛋白与这段 DNA 结合保护了其中的 G 和 A 不被甲基化。第二个技术是用能在磷酸上加上一个乙基的试剂处理 Cro-DNA 复合物，特定的磷酸被乙基化后再定位。仍然是某些磷酸不被乙基化，这也可能是由于 Cro 蛋白的结合使这段 DNA 中的磷酸不被乙基化。前两个技术都是利用 DNA 上与蛋白质作用的特定位点已被 Cro 蛋白保护。第三个技术与前两个不同，该技术是利用在一次位置测定中，通常可以随机改变 1～2 个磷酸，从而抑制 Cro 蛋白的结合。用试剂短暂处理使 DNA 分子中一两个磷酸被随机的乙基化，再加入过量的 Cro 蛋白，测定任一 DNA 分子形成 Cro-DNA 复合物的能力。有些被乙基化的 DNA 分子能与 Cro 蛋白形成复合物，但另一些则不能。测定不能形成复合物的 DNA 分子中被乙基化的位置，这些位置的磷酸被乙基化后干扰了 Cro 蛋白与 DNA 的结合。因为 Cro 蛋白只和没有乙基化的磷酸结合，所以这个磷酸区一定就是接触区。注意这个位置与第二个技术中乙基化的位置不同。目前还没有关于 Cro 蛋白与胞嘧啶和胸腺嘧啶结合的资料。

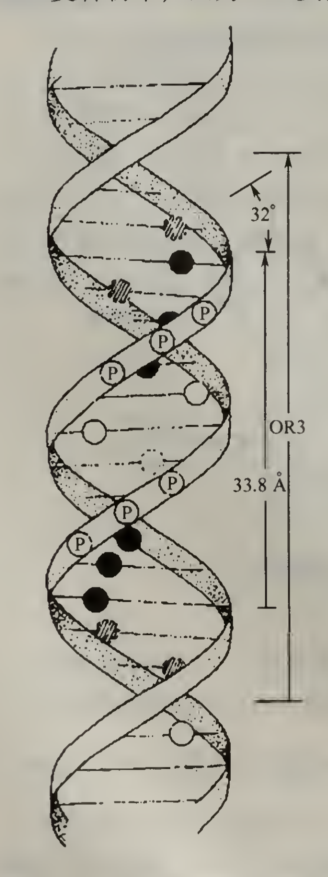


图 2-39　λ DNA 上与 Cro 蛋白接触的 OR3 区的位点

图 2-39 表示 λ DNA 双螺旋中 Cro 蛋白结合的片段。当 Cro 蛋白结合时，G 和 A 不被甲基化，图中还指明乙基化的磷酸阻止蛋白结合甲基化位点（N-3 和 N-7）。

图 2-39 说明了所有的接触点都处在 DNA 双螺旋的一边，在 DNA 的大槽中，从几何学来看是合理的。λCl 蛋白也呈现这种排列，它也能与 λDNA 结合；还有 *E.coli*的 CAP 蛋白也是如此，它结合到*E. coli* DNA 的某一调控位点。这种结合可能是 DNA-蛋白质互相作用的一种共同特点。此外还应注意几点：①Cro 蛋白所结合的 A 和 G 存在于两条链上的；②DNA上的接触点差不多是对称排列的。

对含水结晶的 Cro 蛋白的排列分析指出，此蛋白的列阵有二重对称性，这个列阵如图 2-40 中所示。Cro 蛋白在溶液里有时以二聚体存在，并且很像是在晶体里同样的两个单体分子聚集而成的。这个观察很重要，因为 Cro 蛋白结合的 DNA 的碱基顺序也几乎是对称的，Cro 蛋白的对称排列阵式对于它与一个对称碱基顺序的结合是合理的。图 2-41 表示出了 Cro-DNA 复合物的结构。在图 2-39 中结构的对称性也是显而易见的。

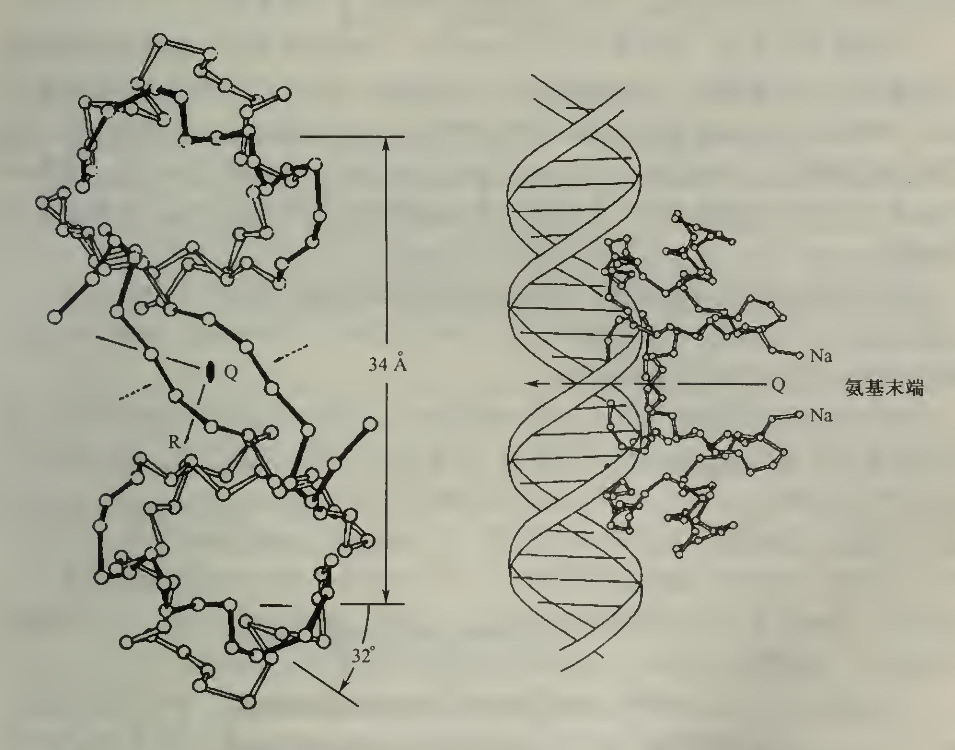


图 2-40　一个蛋白晶体中两个 Cro 蛋白分子的排列

图 2-41　Cro-OR3 复合物的假设结构

仔细观察发现每个 Cro 蛋白单体中的氨基酸以某种方式使其处于能与碱基形成氢键的位置，由此赋予它们结合的专一性；而 Cro 蛋白和磷酸基之间的静电引力则赋予它们结合的强度。

Cro-DNA 复合物是结构最先被弄清楚的序列专一性复合物之一，这些复合物中，结合也发生于特定的碱基序列。大部分被特异性蛋白识别结合的序列都具有某种程度的对称性，而且许多序列特异性的 DNA 结合蛋白是多亚基蛋白。因此我们意识到，蛋白质将会形成对称排列，从而有利于识别对称序列。从其他几种特异的或非特异的 DNA 结合蛋白获得的初步证据表明，这些蛋白质通常在表面有2～3 个伸展的 α 螺旋区。因为每个 Cro 蛋白单体中的一个 α 螺旋置于 DNA 螺旋的大槽中，由此推断，一个 α 螺旋插入一个大槽中可能是 DNA-蛋白质相互作用的一个普遍特征。组蛋白在染色体中的结合也是如此。

4. 生物膜

生物膜（biological membrane）由蛋白质和脂质组成的装配体组建而成。生物膜的结构具有共同的特征，但不同功能的膜存在小小的差别。

生物膜有许多种，每种都有它自己的功能。外膜的最基本的功能就是将细胞内含物质与外环境隔开。然而细胞需从环境中输入养分，因此外膜必须是可透性的。而膜的透性必须是选择性的才能控制细胞内化合物的浓度。获得选择性的透性是通过限制大部分细胞内外物质的自由通过的方法，并且通过分子泵、通道和隘口来允许运输特别的物质。外膜还常常具有信号分子的受体，也具有传递电子的功能。

除外膜外还有许多种内膜，内膜的功能是将细胞内不同的化合物分室以及提供吸附表面使一定的分子进行化学反应。有些内膜包含有酶系统，如植物细胞叶绿体内膜中发生光合反应的酶系统。

细胞膜由双层脂质分子构成，膜中的蛋白质与脂质以疏水作用相互结合。膜蛋白主要是 α 螺旋的球状结构，而不是 β 折叠结构。膜不是静态的，细胞膜中的蛋白质和脂质都能自由运动，这就是流动镶嵌模型（fluid mosaic model）或流动脂质-球蛋白镶嵌模型。这是个动态模型，认为细胞膜由镶嵌着球蛋白分子脂质双分子层按二维排列的液体组成。膜双分子层大部分为轻油般稠度的液状体，具流动性，能在膜平面进行侧向移动而不是上下移动，这样可保持膜的不对称构象的稳定性。脂质双分子层中，内外两层的脂质分子是非对称性的。

有的膜蛋白分子的非极性端嵌入在脂质双分子层的疏水区，极性端则外露于膜的表面；有的膜蛋白则横穿膜层，其两端露于膜的两侧（图 2-42)，它们在膜中的分布也是不对称的。

生物膜的流动性包括膜脂和膜蛋白两类运动状态。膜脂运动有几种方式：①脂酰烃链围绕 C-C 链旋转而导致异构化运动；②磷脂分子围绕与膜平面相垂直的轴左右摆动或作旋转运动；③磷脂分子在脂膜内作侧向扩散或侧向移动；④磷脂分子在脂双层中作翻转运动（f1ip-flop)。膜蛋白的运动可分为侧向扩散和旋转扩散两种。

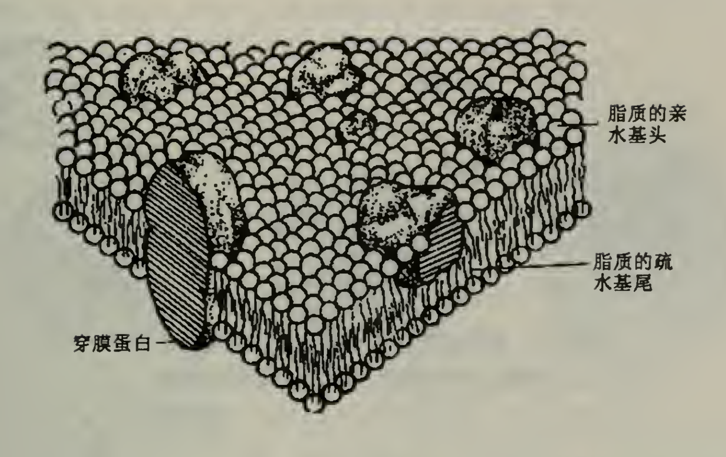


图 2-42　流动镶嵌模型

影响膜脂流动的因素很多，除温度外，还有膜蛋白、遗传因子、pH、离子强度、金属离子、药物作用、膜脂的组成等。其中膜的组成影响较大。

动物细胞由一个结构复杂的膜包围，而细菌、植物和单细胞真核细胞由一个多层结构的细胞壁包围住。细胞膜作为细胞壁的其中一层。细胞壁包含有几个半固态层，在各层之间有众多自由扩散的分子。在实验室里处理细胞，破坏细胞壁中的一层或多层，会引起物质从一个区域到另一个区域的泄漏或从细胞内向细胞外泄漏。

细菌以几种不同的形状存在，每种形状由细胞壁中特定的生物大分子决定。因此对于细胞壁的结构很难作一个准确的叙述。以革兰氏阴性菌为例（如 *E. coli*），其细胞壁的结构如图 2-43 所示。而革兰氏阳性菌和革兰氏阴性菌不同的细胞壁结构如图 2-44 所示。

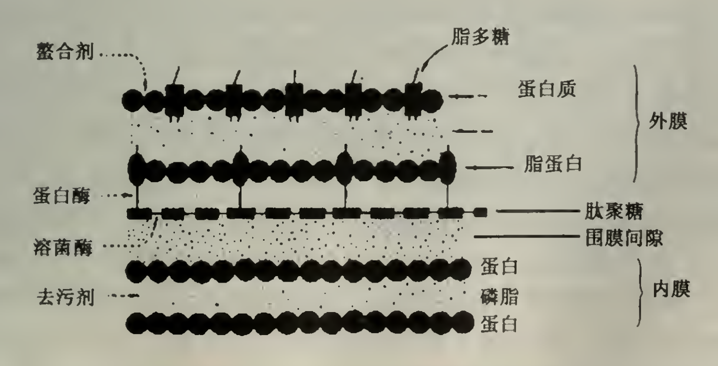


图 2-43　革兰氏阴性菌的细胞壁结构图

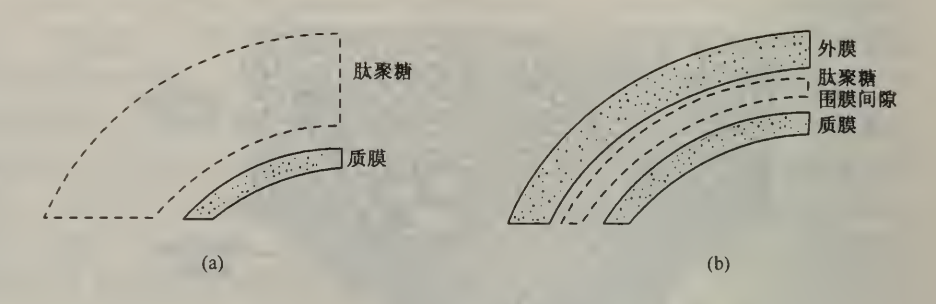


图 2-44　革兰氏阳性菌和革兰氏阴性菌的区别

(a) 革兰氏阳性细胞膜；(b) 革兰氏阴性细胞膜。

5. 复杂聚集物的自我装配

前边已描述了由两种或多种大分子组成的大分子聚集物及其精细结构，但还没谈及它们是如何装配的。复杂聚集物的形成有两种可能性：①由亚基自动形成结构而不需任何外加的能源，这就称为自我装配（self-assembly）；②在酶或一些模板指导下聚集，需要能源。

两种方式装配例子都存在，本节主要介绍一种自我装配的例子——烟草花叶病毒粒子（tobacco mosaic virus，TMV）的自我装配。

烟草花叶病毒是一个长 3000Å，宽 180Å 的棒状病毒。病毒的蛋白质外壳含有 2130 个相同的亚基，这些亚基围绕一个含 6395 核苷酸的单链 RNA 呈螺旋排列形成螺旋对称结构。螺旋壳体中央有直径 40Å 的轴孔，亚基呈右手螺旋排列，螺距为 23Å，每一螺圈上亚基数目为 161/3 个。大约每三个核苷酸结合着一个蛋白质亚基。

TMV 装配的最简单的机制可能是蛋白质亚基逐步加到 RNA 上。但这个机制从逻辑上是不能令人满意的，这样的多步骤过程必定很慢，大约 17 个蛋白亚基必须在这个复合物完成之前靠自身力量加到易弯曲的 RNA 分子上而形成一个螺旋圈，从而获得稳定性。这个问题可以通过许多亚基先形成一个稳定的复合物，然后 RNA 再掺入的方法得到解决。

实验证明，TMV 装配反应的起始成分不是单个的亚基，而是在近生理条件下（pH 7.0，离子强度 0.1～0.3）由 34 个亚基聚集形成的 20S 双盘（double disk）结构。体外实验表明，若缺乏形成双盘结构的条件，壳体蛋白与病毒 RNA 则不能发生结合反应。RNA 嵌入距 20S 双盘中心 40Å 处，RNA 与 20S 双盘结构的结合导致圆盘结构转变为螺旋结构。一旦这种装配起始复合物形成，蛋白质亚基可逐个加入，完成 RNA 的包装，最后成熟为病毒颗粒。

装配反应开始于距病毒 RNA 3′端 830 个核苷酸的区段。此装配起始部位有

一发夹结构（图 2-45）。TMV 自我装配过程是由 RNA 装配起始部位的发夹结构嵌入双盘结构之间、从内侧与双盘结合开始的。

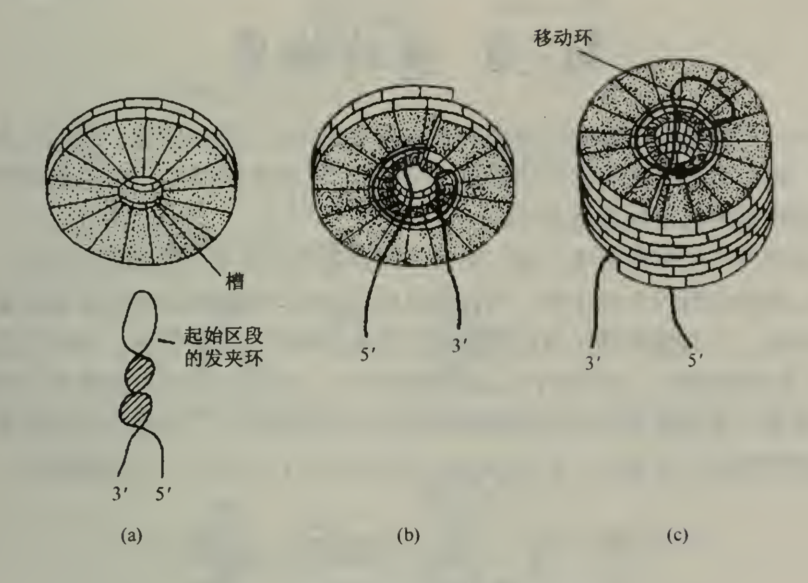


图 2-45 烟草花叶病毒装配的移动环模型

（a）RNA 起始区的发夹插入 20S 圆盘的槽内

（b）圆盘结构转变为垫圈形结构嵌住 RNA

（c）蛋白质先将 5′端装配完成后，转向 3′装配，直到完整颗粒产生。

因装配的起始部位既不是随机的，也不是在 RNA 链的任一末端，而是在 RNA 的内部，所以装配方向必是从内部起始部位双向生长。实验证明装配首先是向 5′端进行，当 5′端的包装完成后，才继续向 3′端延伸。

当 RNA 发夹结构专一性地与外壳蛋白双盘结构结合时，RNA 的两个末端不是分别在装配起始复合物的两边，而是在同一边出现两条链：一条是 5′端的长链，另一条是 3′端的短链。装配开始后，蛋白质亚基加入装配起始复合物，螺旋顶端不断生长。由于蛋白质亚基与 RNA 的结合，使 5′端的长链穿过正在生长的螺旋结构的中心孔，正在装配的颗粒不断向 5′端延伸。在此过程中，装配过程的主要特征是在颗粒的主要生长端有一个移动的 RNA 环（图 2-45）。待 5′端完全拉出后，沿 5′方向的装配完成，然后颗粒转向 3′端生长，直到完整的颗粒产生。

第三章　遗传物质

任何一种生物的遗传物质都携带决定该生物性状的遗传信息。它们从亲代传到子代。绝大多数生物的遗传物质是 DNA。少数细菌噬菌体、许多植物病毒和一些动物病毒的遗传物质是 RNA。

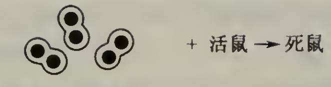
在现代分子生物学诞生之前，很多人认为蛋白质分子可能是遗传物质，因为在当时已知的细胞内大分子中，只有蛋白质在结构和化学性质方面是足够复杂和多种多样的，从而能够满足遗传物质需要变化多端的信息的特点。当时并没有认为 DNA 是遗传物质，只知道 DNA 在细胞核内，并且认为 DNA 太简单不可能作为遗传物质，从而推测 DNA 只是染色体的一些结构成分。有两个实验结果改变了对遗传物质的上述看法，正是因为这两个实验工作才创立了分子遗传学。

第一节　遗传物质的证明

一、转化实验

认为 DNA 是遗传物质的思想是从 1928 年 F.Griffith 的观察开始的，他研究细菌和人类肺炎的关系。所用的细菌是肺炎双球菌，这种细菌的毒力依赖于它们有无荚膜。有荚膜的细菌能抵抗机体对它们的破坏，这种肺炎双球菌在琼脂平板上长成的菌落是光滑型（S 型，smooth edged），用光滑型菌株感染小鼠能使小鼠致死［图 3-1（a)]。后来 Griffith 分离到了一种形成粗糙型菌落的肺炎双球菌（R 型，rough edged）突变株，接着他做了一个很重要的观察。他用加热杀死的 S 型菌注射小鼠，小鼠存活下来。但把加热杀死的 S 型菌和活的 R 型菌混合注射小鼠后，出乎意外地小鼠死了，而且从血液中分离到了活的 S 型菌，因此在加热杀死的 S 型菌中存在一种使活的 R 型菌转变成 S 型菌的因子，他把这种现象称为转化（transformation)。Griffith 解释说是热杀死的 S 型菌的存在导致那些活的 R 型菌恢复了合成荚膜的能力。3 年后这一观点得到了验证，因为只要把 S 型菌加热杀死后加入到体外培养的 R 型菌培养物中，也可使 R 型转化为 S 型。两年后又证明仅把 S 型的细胞提取液加到培养有 R 型菌的培养物中也发生 R→S 的转化（图 3-2)。

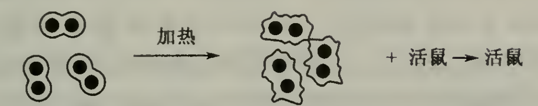
(a) S菌株杀死小白鼠



(b) 无荚膜的R株不杀死小鼠



(c) 加热杀死的S株不杀死小鼠



(d) 分开时都不能杀死小鼠的R株和加热杀死的S株，加在一起后杀死小鼠

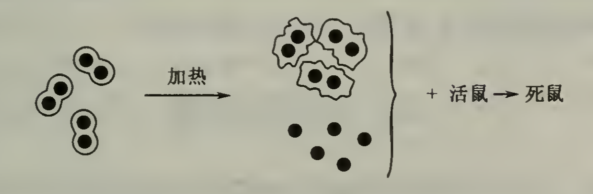


图 3-1　Griffith实验表明用致病菌细胞抽提液处理非致病菌使其转变成致病菌

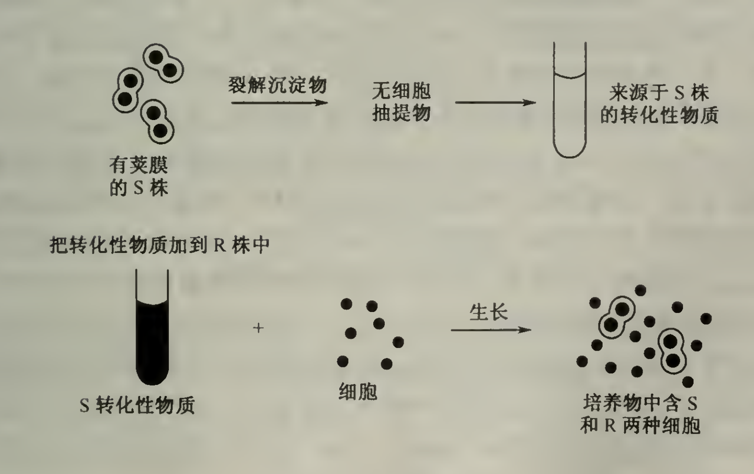


图 3-2　转化实验

为了弄清这种转化因子的化学本质，Oswald Avery，Colin Macleod 和 Maclyn McCarty 对S型细菌细胞提取液进行了分离。他们发现用一系列的化学和酶学方法把提取液中的蛋白质、类酯、多糖、核糖核酸（RNA）去掉后，并不影响R

→S 型的转化。最后他们对提取液进一步纯化，发现只要把纯化的 S 型菌 DNA 以 6×10^{-9}g 的剂量加到 R 型菌细胞培养物中就足以导致 R→S 的转化，因此他们认为这种转化因子就是 DNA。

Avery 等在 1944 年发表这个结论时引起了人们的争论，因为在这之前从未有人考虑过 DNA 是遗传物质，认为 DNA 只是简单的四种核苷酸，与淀粉一样仅是些简单重复的大分子，并且认为蛋白质可能是遗传物质。所以当 Avery 等的结论发表后不仅没有被人们接受，而且受到了人们的反对。反对者认为，无论怎样纯化的 DNA，仍有可能藏留一丝蛋白质，可能就是这一丝蛋白质在转化中起作用。Avery 为了印证他们的结论做了大量工作，他们用各种蛋白水解酶处理提纯的 DNA 都不影响 R→S 的转化，但只要用脱氧核糖核酸酶稍微处理一下 DNA，R→S转化活性就立即消失，这就证实了转化因子的确是 DNA。Avery 还发现：①从转化了的细菌中可以提取到比原来高许多倍的转化因子，说明转化因子具有自我复制能力。②能指导细菌合成多糖，这也正是基因所具备的特点。因此，Avery 的转化实验实际上证明了 DNA 是遗传物质。

二、化学实验

1928 年以前，科学界普遍认为 DNA 只是四种核苷酸组成的物质，人们相信 DNA 是由等物质的量的腺嘌呤（A)、胸腺嘧啶（T)、鸟嘌呤（G）和胞嘧啶（C）组成。这个不正确的结论主要来自两个实验，一个是在 DNA 未溶解彻底和未作 DNA 纯度鉴定之前就进行 DNA 的化学组成分析，这样使定量分析的结果很不准确。第二是当时用于碱基分析的 DNA 一般从真核生物中提取，而真核生物 DNA 的碱基是由四种碱基接近等物质的量组成的。后来 Chargaff 广泛地使用了多种生物的 DNA，并且采用纸层析新技术分离各种碱基，用紫外光吸收测定每种碱基的含量，发现 DNA 的碱基组成很不一样，组成不同种 DNA 的四种碱基的物质的量浓度是可变的。这些结果初步满足了 DNA 作为遗传物质的需要。由于 Chargaff 实验结果的发表，DNA 是简单四核苷酸聚合物的思想很快消失，同时 DNA 就是基因的思想兴起。不久，同一实验室的 Alfred Mirsky 和 HansRis 及另一实验室的 R Vendrely 和 A Boivin 分别发现各种不同的有机体体细胞中的 DNA 是精子中的 2 倍，这正是遗传物质应有的特性。这些发现使经典的染色体遗传学原则得以产生。这个结果发表后，DNA 的遗传物质性质很快变成了流行的概念。

三、Blendor 实验

DNA 的遗传学特性的最有力的证据来自于大肠杆菌 T2 噬菌体实验。由于用

厨房里的搅拌器（blendor）作为实验装置的一个主要部件，所以叫做 Blendor 实验。Alfred Hershey 和 Martha Chase 证实注入到细菌中的噬菌体粒子 DNA 含有合成子代噬菌体所需要的全部遗传信息。

组成单个噬菌体的 DNA 包装在蛋白质外壳内，在噬菌体粒子中只有 DNA 是含磷的物质，蛋白质外壳含氨基肽，仅甲硫氨酸和胱氨酸含硫原子。在^{32}P 标记的培养基中培养能得到^{32}P 标记了 DNA 的噬菌体。如果在^{35}S 标记的培养基中培养则得到^{35}S 标记了蛋白质外壳的噬菌体。用这两种不同标记的噬菌体分别去感染细菌就可以根据其放射活性在被感染的细胞上将噬菌体的 DNA 和蛋白质定位。Herchey 和 Chase 证明了这些噬菌体注入细菌细胞内的是^{32}P 标记物而不是^{35}S 标记物。

第一个实验是把^{35}S 标记的噬菌体与细菌混合，吸附几分钟后，离心除去没有吸附的噬菌体，收集沉淀（噬菌体-细菌混合物），将沉淀悬浮后再一次离心，收集上清液和沉淀，分别测定放射活性。发现 80％的放射活性存在于上清液中，沉淀中只有 20％放射活性。这是由于在这个期间大部分吸附的噬菌体已将其 DNA 注入细菌而蛋白质外壳与细菌解离后进入上清液，但仍有少部分蛋白质外壳还留在细菌细胞壁上（图 3-3）。后来证明沉淀中的 20％放射活性确实是由于

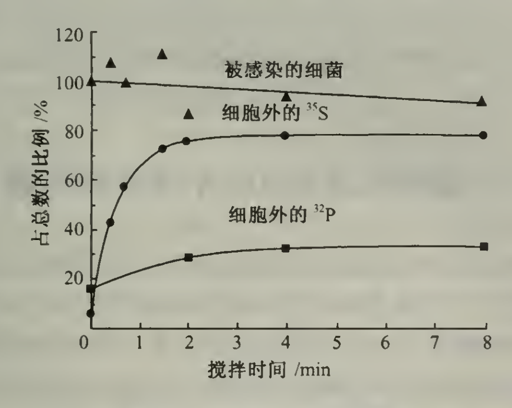


图 3-3　Blendor 实验的数据

噬菌体的尾丝与细菌表面结合太紧，以致 Blendor 也不易将它除去。用^{32}P 标记的噬菌体和细菌混合，用上述方法同样处理后实验结果差别很大，70％的放射活性在沉淀里，30％的放射活性在上清液里。在上清液中的 30％的放射活性可能是由于搅拌细菌时噬菌体破损产生的（几年后发现一些缺损的噬菌体粒子不能将其 DNA 注入到细菌中）。将这些沉淀悬浮在生长培养基中重新保温，能够产生子代噬菌体，这些能产生子代噬菌体的细菌同时可以从亲代噬菌体转移^{32}P 到子代细菌细胞中（图 3-4）。

另一类实验是众所周知的转移实验（transfer experiment），也证明了噬菌体

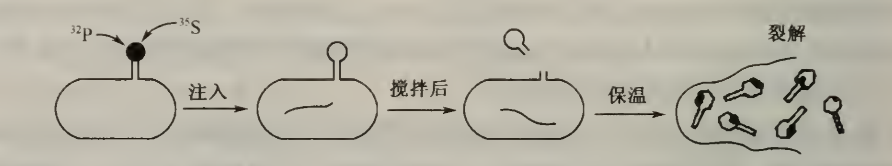


图 3-4　Blendor 实验

的遗传物质含^{32}P而不含^{35}S。在这类实验中，从感染了^{35}S或^{32}P标记的噬菌体的细菌中分离子代噬菌体，对它们进行放射活性分析，结果必然是在子代中发现一些亲代的遗传物质，在子代噬菌体中没有^{35}S，但约有一半注入的^{32}P转移到了子代噬菌体中，这个结果说明，虽然^{35}S可能存在于噬菌体感染的细菌复合物上，但它不是噬菌体遗传物质的一部分。仅有一半^{32}P被转移的解释是：子代 DNA 是随机包装到蛋白质外壳中的，而且并不是所有的子代噬菌体 DNA 都能被成功地包装。

第二节　遗传物质的性质

作为遗传物质首先必须具有稳定性，其次必须具有多样性，还必须具有变异性。

一、遗传信息由 DNA 储存和传递

一个细胞中 DNA 携带和传递的遗传信息包括以下几类：①细胞合成的每一种蛋白质的氨基酸顺序。②合成每种蛋白质的起始和终止信号。③决定在某一单位时间里合成哪些特定的蛋白质并且需合成多少蛋白质分子的信号装置。

这些信息包含在 DNA 分子的碱基顺序中，但这些氨基酸顺序和起始、终止信号并不是从 DNA 中直接获得，而是从 RNA 中间物中获得。DNA 为信使 RNA（mRNA）提供模板，mRNA 的顺序与 DNA 中的一条链互补，蛋白质中的氨基酸顺序根据 mRNA 中的碱基顺序翻译。碱基顺序以三个碱基为一个阅读单位，叫三联密码，每一个三联密码对应一个特定的氨基酸或起始、终止信号。这种分两步走的过程具有重要作用：虽然，蛋白质的合成是连续进行的，但这样分两步走 DNA 分子既不会被经常地使用，也不会直接地进入蛋白质合成机器。通过碱基配对，DNA 上的顺序可以很容易地转录到 RNA 分子上。在酶学系统中，仅仅通过互补形成氢键就可以合成 RNA。这一点在转录一章将详细讨论。DNA 的局部变性有利于碱基的识别过程，当碱基对暂时打开时，RNA 聚合酶能够穿透到 DNA 螺旋中阅读碱基顺序，这对于 RNA 的合成是必需的。

一些起调控作用的蛋白质分子也能识别特异的碱基序列，决定蛋白质是否合成。这些调控蛋白可直接结合于由六个以上碱基对组成的序列上，而且并不需碱基对分开，与结合到 DNA 分子的外槽内相似，这些调控蛋白能识别特异的电子构型并通过电荷、范德华力、疏水作用及氢键与 DNA 序列结合，这种调控蛋白与碱基序列结合的例子有：λ DNA 与 Cro 蛋白的结合、抑制蛋白如 LacI 蛋白以及 cAMP 受体蛋白等与特异碱基顺序的结合。

二、遗传信息从亲代传递到子代

当细胞分裂时，每个子代细胞必须含有同样的遗传信息。也就是说 DNA 分子必须变成两个同样的分子，并且每个分子都携带亲代细胞中所含有的遗传信息，这种二倍化过程叫复制。复制只通过碱基互补形成氢键来进行。

三、携带遗传信息的 DNA 的化学稳定性

在生物进化过程中，一种简单的 DNA 分子至少经历了几千年或更长的时间，在这数百万年中这些含有遗传信息的分子在细胞繁殖的传递过程中只有小小的改变，因而 DNA 分子必须有很高的稳定性。

DNA 分子的脱氧核糖-磷酸骨架是极端稳定的。脱氧核糖中的 C-C 键可以抵抗所有条件下的化学攻击，甚至能抵抗高温下的强酸作用。虽然其磷酸二酯键的稳定性较小，在室温 pH 2 条件下可水解，但这不是生理条件。由于磷酸二酯键的稳定性较小，所以 DNA 的戊糖是 2′-脱氧核糖而不是核糖。在碱性 pH 条件下，RNA 可很快水解成自由的单核苷酸，甚至在 pH 8 时 RNA 分子的磷酸二酯键就打开而形成大小不等的片段，该反应叫做 β-消除反应，它需要糖环上有羟基，除了末端核苷酸上的 3′-OH 外，还需要在 RNA 骨架中有 2′-OH。DNA 中的 2′-脱氧核糖中没有自由-OH，也就是说，DNA 不可能通过这个途径水解。在室温 pH 7～9 条件下检测不出 DNA 的水解。碱基在化学结构上的改变意味着遗传信息的丢失。在一个细胞里有大量的化合物能攻击自由碱基，由此就可以看出 DNA 双螺旋结构的重要意义。DNA 是两条链，这对携带遗传信息来说是过剩的分子，DNA 分子的一条链与另一条链互补，细胞里存在修复系统，除去改变了的碱基后可以通过阅读互补链上的顺序将正确的碱基替换上去，DNA 双螺旋更重要的意义在于双螺旋性质为碱基抵抗化学攻击提供了保护，碱基使疏水环上有带电基团，正是这些带电基团需要保护。这些碱基基团的每对之间的氢键为其提供了第一道防线，碱基的疏水性使碱基堆积，大大减少了碱基被化学攻击而受伤的可能性。氢键位于碱基堆积成的圆柱体的中心，碱基堆积的结果使水完全从碱基排列的方阵中排除出去，由于任何潜在的有害化合物通常通过溶于水产生毒害

作用，所以这些分子很难进入“干”的碱基堆积中靠近和接触碱基，使这些化合物攻击碱基的可能性很小。糖苷键位于DNA的碱基和糖的1′-碳原子处，除了极端pH条件下，糖苷键很稳定，如果碱基不是环状，那么糖苷键的稳定性将降低，嘌呤环上7位氮原子的甲基化极大地降低了糖苷键的化学稳定性，这可能是高度诱变和高度毁灭性的作用模式。氮介子和硫介子曾在第一次世界大战中使用，任何一种介子（氮介子、硫介子等）都提供烷基基团，与鸟嘌呤7位氮原子接触，使鸟嘌呤从糖环上脱落，这样糖环上在1位碳原子上剩下一个—OH，接着磷酸二酯键发生水解。

除胞嘧啶外，碱基本身是很稳定的，胞嘧啶以很低的频率脱氨基形成尿嘧啶，脱氨基作用是个灾难性的变化，因为脱氨基后变成尿嘧啶，而尿嘧啶与腺嘌呤的配对机会比它与鸟嘌呤的配对机会多，这样会引起两个效应：①不正确的碱基将出现在mRNA中，②在DNA复制时A代替G将合成一条新的DNA链（图3-5）。在复制一章中将讨论到在细胞内DNA复制时将用T代替U，这就消除了利用脱氨基作用形成的尿嘧啶的可能性，这就很好地解释了为什么DNA复制中利用T与A配对而不利用U与A配对。如果U是DNA中的正常碱基，就没有办法鉴别是正常的U还是C通过脱氨基作用形成的不正确的U。细胞遵循一个原则：RNA利用U而不利用T，DNA利用T而不利用其他碱基，但产生这种现象的原因并不清楚。这可能是进化中的一个偶然事件，很可能原始的RNA和DNA都只含有胞嘧啶和尿嘧啶，当时只有这两种嘧啶可以利用。后来细胞获得了把尿嘧啶甲基化形成胸腺嘧啶的能力，因此给细胞提供了一个消除引起C→U转变的标准。胸腺嘧啶核苷酸合成的最后一步是脱氧尿嘧啶核苷酸的甲基化，这是重要的一步。胸腺嘧啶转变成三磷酸核苷酸才能掺入到DNA中。

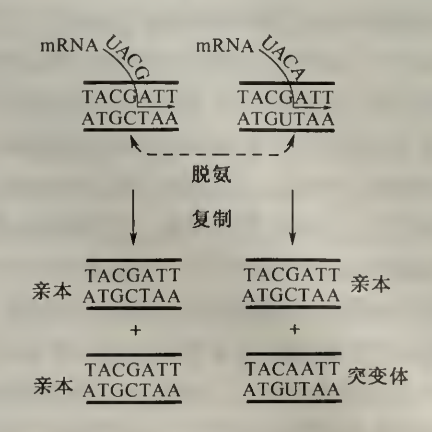


图3-5　在mRNA分子和两个子代DNA分子中胞嘧啶脱氨基形成尿嘧啶

四、遗传物质的变异性（突变）

单个细胞的所有遗传信息都存在于它的 DNA 中，而细胞要能够进化，其 DNA 碱基顺序必须是可变的。并且这种变化后的 DNA 必须是持久的，能传递给子代，这样子代细胞才能具有这种新 DNA 的性质。这种碱基顺序的改变过程叫突变，突变有两种重要的机制：①一个碱基的化学变化给于这个碱基新的氢键性质，使新复制的子代分子中出现一个新的碱基。②复制错误使一个不正确的或一个额外的碱基意外地插入到子代 DNA 分子中去。

一般来说，突变是有害的，并且可能导致细胞死亡。所以突变的比率必须很低或受到控制。防止和控制突变发生通过两种途径完成，第一，DNA 分子无水的疏水核降低了 DNA 对化学攻击的敏感性，这一点如前所述；第二，细胞已经进化了几种修复机制，使变化的碱基或在复制中出现的错误及时纠正。这些修复体系虽然不是完全有效的，它允许以很低的频率产生突变，但是这种很低频率的突变对于生命长期的延续是有用的。不过正如前面所说的那样，这种突变一般来说是有害的，所以突变时不丧失亲代的遗传信息是很重要的。生物可以从两方面防止亲代遗传信息的丧失：①一个种中的其他成员保留亲代的碱基顺序，②双链分子中的一条链是冗余的。通常 DNA 双链中只有一条链发生改变，而 DNA 复制时是以两条亲代链为模板各自合成一条新链，细胞分裂后每个子代细胞里都只含有亲代 DNA 的一条链，这样可以形成一个含有正常 DNA 分子的子代细胞，而另一个子代细胞的 DNA 是突变的。除非这个细胞具有亲代的遗传信息，否则这个突变体可能不能存活。但是如果这个突变体在生存和繁殖方面比亲代生物或这个种的其他任何成员都优越的话，通过很多世代之后，根据达尔文的适者生存

图 3-6 亚氨基形式的腺嘌呤和胞嘧啶及烯醇式的胸腺嘧啶和鸟嘌呤之间的配对

的原理终将导致自然界中突变型代替亲代的野生型。

突变的第二种机制——复制错误，这是由于碱基A和T的异构化性质形成的。A的亚氨基可以与C形成稳定的氢键，而烯醇式的T可与G形成配对（图3-6），产生这种现象的原因可能是在原始的海洋里，核酸分子也有自发出现的其他碱基，可以变构的碱基被保存下来，因为它们允许突变，使生物能适应环境的改变以及适应进化的要求。

第三节　遗传物质——RNA

前一节我们已经讨论了DNA是遗传物质，但这只是一般性的结果，还遗留一些问题必须解答。

我们知道有的动、植物病毒及有些噬菌体不含DNA，只有RNA。DNA作为合成蛋白质的遗传信息的携带者，并且在DNA到蛋白质的合成途径中总是以DNA为模板合成RNA中间物，即中心法则，DNA→RNA→蛋白质。而一个没有DNA的生物只能利用RNA的碱基顺序储存遗传信息作为合成蛋白质的模板。但RNA不能像DNA→RNA→蛋白质系统中那样有效地执行这两种功能，那么RNA作为遗传物质是如何行使其功能呢？下面我们将讨论这个问题。

在前一节我们了解了DNA可以作为遗传物质的优越条件，然而在有些生物中它们缺少起基本保护作用的双螺旋，并且已知有一些病毒和一种噬菌体是双链RNA。在细胞生物中，RNA是不能作为遗传物质的。因为细胞里有许多可对RNA产生化学攻击的物质降低了RNA在细胞中的化学稳定性。唯独在噬菌体和病毒里RNA可以作为遗传物质，因为它们包装在一个不溶解的蛋白质外壳里，使它们在环境中受到保护。在宿主中RNA分子大部分时间以不活动的粒子状态存在，并以很低的频率复制，当繁殖时，复制非常迅速，在很短时间里产生大量的子代粒子，以其绝对的数目来补偿较低化学稳定性造成的RNA的不足。这样RNA噬菌体和RNA病毒逐渐形成补偿特性而得以存活。

第四节　可转移的遗传因子

病毒、质粒和转座因子都是染色体外的可以移动的遗传物质，它们都必须依赖于宿主而复制。这一节我们主要介绍：质粒的一般性质、复制机制以及几种原核和真核细胞中的质粒；转座因子的特征及转座机制；从进化角度讨论病毒、质粒及转座因子的关系。

一、质　　粒

质粒（plasmid）是在许多细菌和某些真核生物中发现的染色体外的环状DNA分子。在正常情况下，质粒可以随宿主细胞的分裂而分配到子代细胞中去，但在细胞分裂过程中有时也产生无质粒的细胞，而且这种细胞也可以存活，然而许多质粒含有的质粒基因在某种条件下对于宿主的生长是必需的。例如R质粒携带抗抗生素的基因，含有这种质粒的细胞可在某种抗生素存在的环境中存活；有些质粒如F′，携带来源于宿主染色体上的几个基因，例如携带基因 leu^{+}，这样的质粒对于 leu^{-} 的细胞是必需的。还有许多其他类型的质粒常常可以通过它们所携带的基因来测定。

1. 质粒的一般性质及类型

虽然不同的质粒具有互不相同的基因和调控位点，但它也像噬菌体一样，主要依赖于宿主的代谢功能进行复制，例如质粒可以正常地利用宿主的大部分复制机器。然而每一种质粒都有控制复制时间和控制在每个细胞中拷贝数的基因（严紧型质粒每个细胞只有1～2个拷贝，松弛型质粒每个细胞有10～100个拷贝）。在细胞分裂时需小心地控制分配到子代细胞中的质粒个数，即假如一个细胞内的质粒已为细胞分裂准备了两个质粒分子，那么每个子代细胞中就应该有一个质粒DNA分子，如果这一过程出现差错，分裂时就会出现无质粒细胞，这种情况出现的频率是 10^{-4}/代。质粒的这些稳定性质受质粒和染色体的联合控制，但联合控制的机制还不十分清楚。

在不同的大肠杆菌菌株中发现了许多类型的质粒，已研究得较多的有三种，F、R和Col质粒。这些质粒虽然具有一些共同的性质，但也有许多不同的地方。①F性因子，能将染色体基因和它本身转移到无质粒的细胞中。②R抗药性因子，抗一种或多种抗生素并且能将抗性转移到无R质粒的细胞中。③Col大肠杆菌素因子，能合成大肠杆菌素，大肠杆菌素可以杀死亲缘关系很近但不含Col质粒的菌株。

除酵母killer质粒为RNA分子以外，已知的所有质粒都是环状超螺旋DNA分子，它们的分子质量从 10^6Da（小质粒）到 10^8Da（大质粒）不等。表3-1列出了目前研究得比较多的几种质粒的分子质量。根据质粒的超螺旋性质，通过离心、密度梯度离心和电泳很容易将其和染色体DNA分离。

通常分离完整DNA都要经过去蛋白步骤，但从大肠杆菌中分离质粒DNA一般没有这个步骤，大约有一半的超螺旋质粒DNA分子中含有三个紧密结合的蛋白质分子，这种DNA-蛋白质复合物称作松弛复合物（relaxation complex）。如果用加热、碱、蛋白水解酶或去污剂处理这种复合物，因这三种结合蛋白中有一

种具有核酸酶活性，可使质粒DNA的一条链上产生切口，使超螺旋松弛，形成环状切口形式。这个切口在DNA上的一个专一位点上产生。在超螺旋松弛过程中，两个小蛋白质分子释放，而一个大蛋白质分子仍共价结合在切口DNA的5′末端。松弛酶只能使超螺旋DNA分子专一位点上产生切口，而不能使松弛环产生切口。产生切口在质粒由一细胞转移到另一细胞的接合转移中起作用。

表3-1　几种质粒及各自的性质

质粒	分子质量/ $\times 10^6$Da	拷贝数	自我转移能力	表型特征
Col				
Col E1	4.2	10～15	不	大肠杆菌素E1（膜改变）
Col E2	5.0	10～15	不	大肠杆菌素E2（Dnase）
Col E3	5.0	10～15	不	大肠杆菌素E3（核糖体RNase）
性质粒				
F	62	1～2	能	性纤毛
F′lac	95	1～2	能	性纤毛：lac操纵子
R质粒				
R100	70	1～2	能	Cam^r　str^r　Sul^r　Tet^r
R64	78		能	Tet^r　Str^r
R6K	25	12	能	Amp^r　Str^r
PSC101	5.8	1～2	不	Ter^r
噬菌体质粒				
λdv	4.2	≈50	不	λ基因 *cro*，*CI*，*O*，*P*
重组质粒				
pDM500	9.8	≈20	不	果蝇组蛋白基因
pBR322	2.9	≈20	不	高拷贝数
pBR345	0.7	≈20	不	ColE1型复制
酵母质粒				
2μm	4.0	≈60	不	不知道其基因

2. 质粒的复制机制及其拷贝数的控制

从表3-1中可以看到有些质粒有转移能力，而有些质粒无转移能力。一般分子质量小的质粒无转移能力，而分子质量大的质粒有转移能力，如分子质量在10×10^6U以下的质粒一般无转移能力，而分子质量在25×10^6U以上的质粒一般具有转移能力，如F性因子和R抗药性因子。

在遗传结构、行为、分子性状方面研究得最早和最为详细的质粒是大肠杆菌的性因子（F），具有这种质粒的细菌能形成表面毛状物，通过表面毛状物介导，

细菌在与其他细菌接触时可发生接合作用，并将质粒 DNA 传递给对方。在 F 因子的基因组中，具有一些容易与其他 DNA 发生重组的特殊结构，常和寄主染色体发生重组，在接合传递时把一部分染色体转入受体菌内。

根据 F 因子的有无和存在状态，可以把大肠杆菌区分为三种类型，即 F^-：细胞中无 F 因子；F^+：细胞中有 F 因子，存在于细胞质中；Hfr：细胞中有 F 因子，整合在染色体上，即所谓的高频重组（high frequency recombination）菌株。F′因子是 F 因子上带有寄主染色体基因，这是由于 Hfr 错误割离造成的。

质粒的转移过程可分为四步：①有效接触。供-受体对的形成。②动员。为质粒转移做准备。③DNA 转移。④受体中形成一个有复制功能的质粒（repliconation）。

许多质粒的转移并不具备上述全部过程，所以有几个名词需要说明。

接合质粒是携带有有效接触功能基因的质粒。可动员质粒是能够为转移准备自己的 DNA 的质粒。自身转移性质粒是具备接合和动员两种功能的质粒，如 F 因子。图 3-7是 F 因子转移的详细过程。

当质粒的松弛复合物中的切口蛋白在质粒 DNA 的一条链的特定顺序上产生切口时（这个特定顺序叫转移原点），动员开始。环状 DNA 从切点处开始滚环复制（切口蛋白仍结合在 5′末端并且转移到受体细胞），环化后再复制形成环状超螺旋质粒 DNA 分子。

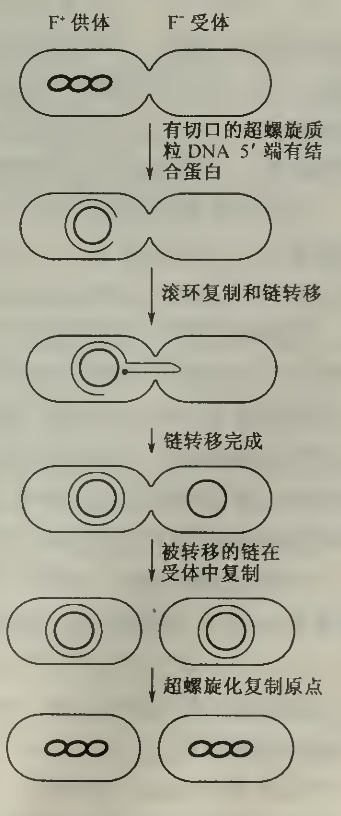


图 3-7　F 因子的转移过程

在接合转移中，供体和受体中都发生了 DNA 的合成。供体中的 DNA 合成叫做供体结合 DNA 合成，受体中的 DNA 合成叫受体结合 DNA 合成。以上所说的质粒 DNA 的转移指的是细胞与细胞接触后的转移。在实验室中把自由质粒 DNA 和适当处理的受体细胞混合，质粒 DNA 也可以转移，这叫做转化。如果细菌用冷 $CaCl_2$ 溶液处理后就获得了吸收 DNA 分子的能力。

我们知道质粒 DNA 的复制在很大程度上依赖于寄主的复制机器，但不同的质粒对寄主复制机器的利用程度不同，例如小型质粒 ColE 1 可以在体外细菌提取物中复制而不依赖质粒基因本身编码的蛋白质（当然不能排除质粒蛋白质在体内对于质粒的复制具有必要性的可能）。另一些质粒如 F 的温度敏感突变体在 42℃时质粒不能复制，含这种突变体质粒的细

菌在42℃培养几代之后，许多子代细胞中不再含有质粒。许多质粒的复制不需要寄主与复制有关的全部基因产物，例如F因子可以在 *E.coli* dnaA（Ts）菌株中在高温下复制，但染色体DNA则不复制，这说明F因子复制不需要寄主的 *dnaA* 基因产物。另一些质粒复制需要寄主基因产物，这个产物对寄主也是需要的，如细菌突变株中 polI 活性降低使细菌不能正常生长，也不能支持 ColE 1 的复制。

不同的质粒的复制机制也不同：

1）复制的方向纯粹地单向复制和纯粹地双向复制在质粒的复制中都曾观察到。还有些质粒存在这两种复制方式。例如质粒 RK6 先按一个方向复制然后在同一复制起始点以相反的方向复制后边的部分。

2）终止按单方向复制的环状DNA分子，复制终点就在复制的起点。双向复制的环状DNA分子的复制终点有两种类型，一种是复制叉延伸到同一区域时终止，例如λ噬菌体。另一种是在固定的位置终止，常常是一个复制叉先到这一位点终止，另一复制叉后到达这一位点，目前对这种固定终止的信号已有所了解。

3）复制形式。在对复制研究得清楚的质粒中，复制以 butterfly 模式进行，在复制的部位解旋，不复制的部位仍为超螺旋，复制时的形状像θ，所以称为θ复制。当一次复制完成时，其中一个环必须被切开（可能是 DNA gyrase 作用），复制后产生一个切口环状分子和一个超螺旋分子，切口环状分子随后再被连接形成超螺旋。究竟这是不是质粒复制的普遍机制，目前还不清楚。

质粒拷贝数的控制。我们知道有些质粒是低拷贝（每细胞1～2拷贝），有些则是多拷贝（每细胞10～100拷贝）。那么各种质粒在细胞中的拷贝数是如何控制的呢？有人提出由质粒基因编码一种抑制物抑制质粒的复制，质粒进入寄主后，在质粒基因没有表达之前，质粒可以利用寄主复制机器进行复制，复制一到两次后，质粒基因产物（抑制物）表达抑制质粒的复制。这样细胞分裂后每个子代细胞中只有1～2个质粒，这种低拷贝数质粒可能只需受抑制物单体分子作用，所以对复制的抑制不需要高浓度抑制物的存在。而对多拷贝质粒来说，其复制的抑制受多聚体抑制蛋白的作用，因此需要抑制物达到高浓度时才能抑制其复制，因此这种质粒有时间复制多次，产生高拷贝。

在杂合质粒 pSC 134 中可以看到质粒拷贝数的控制情况。pSC 134 是由完整的质粒 ColE 1 和质粒 pSC 101 杂合组建的。这两种质粒的拷贝数分别是18和6个，从 pSC 134 中可以得到以下三个事实（图3-8）。

①质粒 pSC 134 从 ColE 1 的复制原点进行复制具有16个拷贝，差不多等于 ColE 1 的拷贝数。

②如果把 pSC 134 放到一个 $polA^-$ 细胞中，ColE 1 原点不能开始复制，pSC 101的原点可以用来复制，pSC 134 在 $polA^-$ 细胞中的拷贝数是6，等于

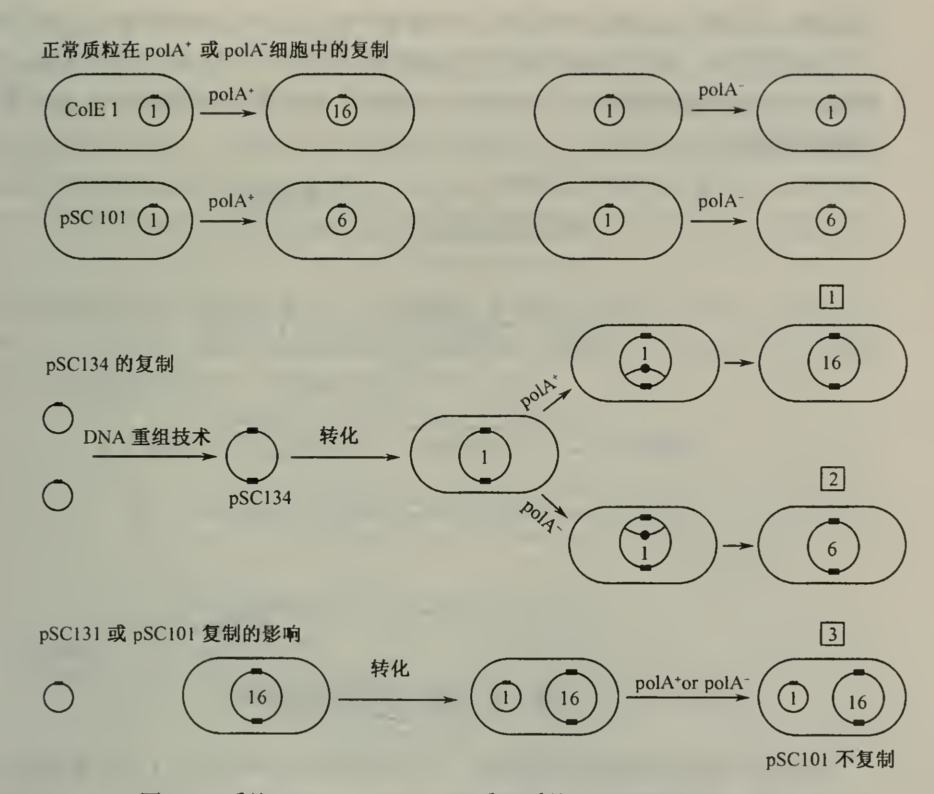


图 3-8　质粒 ColE 1、pSC 101 和重组质粒 pSC 134
在不同细胞中的拷贝数
图中数字（1、6、16）指拷贝数。

pSC 101的拷贝数，这两个结果说明拷贝数的控制与复制原点的利用有关系。

③如果把 pSC 101 引入到一个含有 pSC 134 质粒的细菌中，pSC 101 则不能复制，这个结果表明结合在 pSC 101 上的抑制物是由 pSC 134 质粒制造的。

关于抑制物模型有很重要的一点必须明白，因为在一个特定的细胞中同种质粒的各个分子是完全相同的，抑制物不能分辨是这个分子还是那个分子，当有活性的抑制物分子的浓度比较低，不足以使所有的质粒 DNA 分子的复制受到抑制时，抑制物随机地和少数 DNA 分子结合抑制复制。如果一个质粒复制成两个子代质粒，这两个分开的子代质粒也和其他质粒一样具有复制能力，因而实际上可被复制的分子也是从整个质粒群体中随机挑选的。

对质粒 ColE 1 的抑制机制已十分清楚（图 3-9）。ColE 1 合成的起始过程是从复制原点上游 651bp 处转录一个长的 RNA（叫 RNAII）开始的，这个 RNA 分子的合成也像典型的 RNA 聚合过程一样，只有 3′端少数几个碱基与 DNA 保持配对状态。由于分子内部的互补，释放的 RNA 形成具有几个茎环的结构，这些

形成的二级结构以某种方式使新产生的 RNA 与 DNA 保持接触（从转录起始点 555 碱基开始），大约 100 个碱基与 DNA 形成 RNA-DNA 杂合区，RNase H 在复制原点处的 RNA 链中产生一个切口，产生 3′-OH 末端，确保在 3′-OH 末端加上脱氧核苷酸合成 DNA。

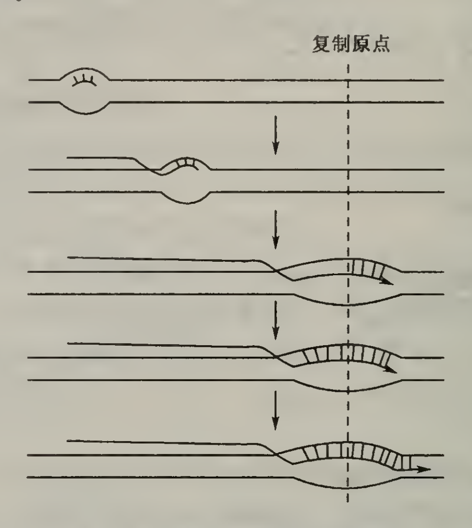


图 3-9　质粒 ColE1 复制的起始

另一个 RNA 也在原点区转录，这个 RNA 分子叫 RNAI，在距 RNAII 转录起始点右边 110 碱基处从右向左转录（图 3-9 中的方向）。RNAI 有 108 碱基长度并且左向转录，也就是说 RNAI 的 DNA 模板链与 RNAII 的模板链互补，即 RNAI 与 RNAII 的 5′端互补，RNAI 可以与 RNAII 形成双链结构以防止 RNAII 的二级结构形成，从而影响 RNAII 与 DNA 形成杂合双链而使 RNase H 不能在复制原点产生切口，这样也就抑制了 ColE 1 的复制从而控制拷贝数。

另一调控因子——质粒编码的一种小的蛋白质 Rom 也是很重要的，但作用方式还不清楚，Rom 决定 RNAI-RNAII 双股分子形成的速度。如果 Rom 蛋白的浓度太低，双链 RNA 不能形成，RNAII 可以形成茎环构象，RNase H 可以切割 RNAII，DNA 的合成起始。

高拷贝数质粒可以表现出扩增现象，如果把氯霉素或别的蛋白质合成抑制剂加到含质粒的细菌培养基里，细菌染色体的合成被抑制，但质粒 DNA 的合成则不受抑制，每个细菌中质粒的数目可增至 1000 个以上。这是因为与细菌 DNA 合成有关的一类起始蛋白（如 DnaA 蛋白）不能合成，但如果质粒复制只是利用细菌的稳定的复制蛋白（诸如 PolI、PolIII、RNase H、RNA 聚合酶）或利用质粒编码的一些稳定的蛋白质，质粒 DNA 的合成就会继续下去。事实上，控制质粒拷贝数的控制因子（例如 Rom 蛋白）是浓度依赖性抑制物，在没有蛋白质合成的情况下，抑制蛋白的数量不足，质粒 DNA 的复制将会反复地起始而不受控

制。由于没有细菌 DNA 的合成，用于质粒 DNA 复制的细菌的复制蛋白效力增加，加上不稳定的 RNA 分子如 RNAI 被降解，形成了质粒 DNA 的过量合成。

关系相近的两种质粒不能在同一细胞中稳定共存，这叫做不相容性。抑制物模型可以解释这种现象。两种完全不相同的质粒如 F 和 ColE 1，它们的抑制物不同，所以两种质粒的复制互不依赖，因为一种类型质粒的抑制物（如 F 的）不能对另一种类型的质粒（如 ColE 1）的复制进行控制，F 和 ColE 1 是相容性质粒。

这种质粒相容性的情况在图 3-10 中的质粒 A 和 B 中则十分不同。A 和 B 的抑制物可以互为调控，而且抑制物的作用是随机的，所以分裂几代之后，就有可能产生同一种质粒存在于同一细胞的情况（图 3-10）。

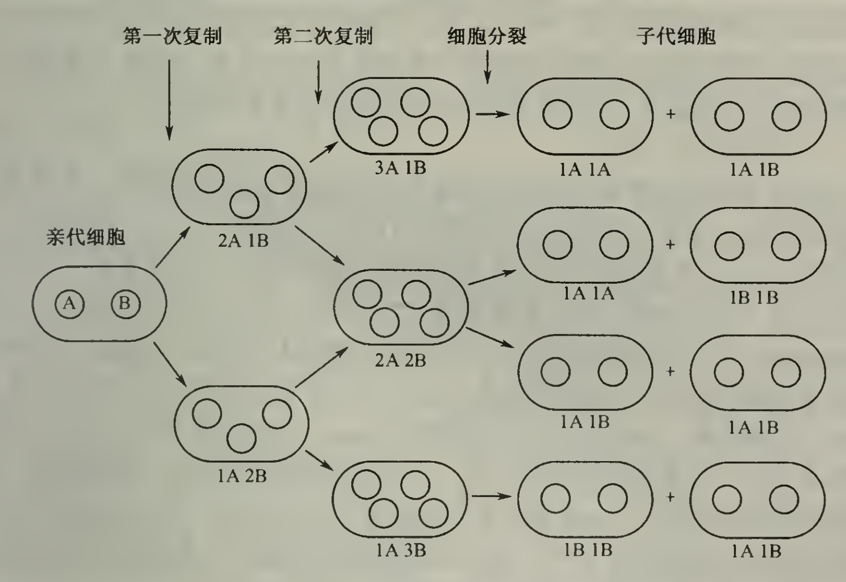


图 3-10　不相容性质粒 A 和 B 在细胞分裂过程中可能的分配情况

这样看来质粒的不相容性产生的原因有两点，①两种质粒具有类似的抑制物，②质粒 DNA 的复制是被随机选择的。

3. 几种质粒

（1）性质粒 F 和它的衍生物

F 是分子质量为 62.5×10^6U（94.5kb）的环状 DNA 分子。利用分子生物学手段已测出了 F 因子的遗传图和物理图（图 3-11），从遗传图上可以看到质粒 F 除结构基因外还有插入顺序 IS2，1S3，γδ，这些

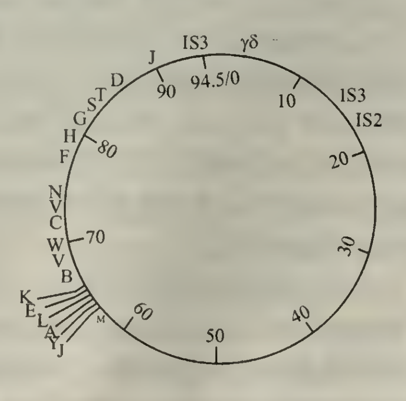


图 3-11　F 因子的遗传图

都是可转座的因子，它们的性质将在后面讨论。

F的一个重要性质是它能整合到细菌染色体上形成Hfr，整合过程是通过互相交换，就像λ噬菌体形成溶源菌一样。但F与细菌染色体的整合方式与λ噬菌体的方式有以下几点不同：①F上的交换位点不是单一的，虽然交换主要发生在IS3即物理图上94.5的位置，但交换也发生在其他位置上的IS3和γδ处。②染色体上有许多位点能够发生整合，现在已知的就不少于20个，但F对每个位点的亲和力不同。③F因子可以以顺时针和反时针两个方向整合，F因子转移到受体细胞中去是从一个原点以一个方向进行的，有些Hfr菌株可以从两个方向转移染色体。事实上，F因子在一个位点上能以两个方向整合的例子很少。

F因子也可以割离，但是很罕见。这个过程很像λ噬菌体从染色体上割离发生错误而产生转导型噬菌体一样，F因子割离时发生错误后可产生F′因子。

F′质粒在最初的分子遗传学研究中是非常有用的工具，如用F′做部分二倍体用来分析lac操纵子。它们在研究DNA转移中也有重要作用。

(2) 抗药性质粒（R）

抗药性R质粒的明显特征是它能使宿主细胞具有抗抗生素能力并且具有自身转移的能力。许多R质粒由两个邻接的DNA片段组成。一个片段叫RTF（抗性转移因子），它携带着控制DNA复制和拷贝数的基因，转移基因，有时还携带抗四环素（tet）基因，其分子质量为11×10^6Da。另一个片段叫抗性定子（r determinatant），其大小很不同（从几百万道尔顿到大于100×10^6Da），携带另外一些抗生素抗性基因，如抗青霉素（Pen）、抗氨苄青霉素（Amp）、抗氯霉素（Cam）、抗链霉素（Str）以及抗卡那霉素（Kam）和抗磺胺药物（Sul），这些抗性基因通常以一个或多个结合在一起。一些小的抗性质粒失去了转移能力，但仍有抗四环素抗性基因，如pSC101分子质量为5.8×10^6Da，常用于基因工程中。双组分R因子使人想到F′质粒，但R质粒的抗药性基因不是由于RTF的整合和错误割离获得的，而是从转座过程中获得的。

(3) 产生大肠杆菌素的质粒（Col质粒）

Col质粒是可产生大肠杆菌素的大肠杆菌质粒，这种蛋白质能遏制不含这种质粒的大肠杆菌菌株的生长，大肠杆菌素有许多种类，以后边的字母来区别不同的种类（表3-2），每一种大肠杆菌素都以特有的方式抑制敏感细胞。对大肠杆菌素的产生的监测类似于对噬菌体的监测分析，将产生大肠杆菌素的细胞和敏感细胞一起铺平板，在产生大肠杆菌素的菌落周围形成敏感菌不能生长的亮区域，称为抑菌圈或空斑。

大肠杆菌素可能有两种类型——真正的大肠杆菌素和缺陷的噬菌体粒子，后一种的推论来自于对许多提纯的大肠杆菌素的研究。当用电镜检查时只有很少的大肠杆菌素是简单的蛋白质，其他的类似于噬菌体的尾，因而认为是残留的古老噬菌体转录的基因产物。这个假设认为原噬菌体经反复突变失去了复制头部蛋白

和裂解有关的基因，但还完整地保留了编码抑制系统和尾部的基因，大概这些噬菌体负责吸附但不注入 DNA 而使生物大分子合成受到抑制。

表 3-2　几种大肠杆菌素的性质

大肠杆菌素	大肠杆菌素的作用
大肠杆菌素 B 大肠杆菌素 b	损坏细胞质膜
大肠杆菌素 E1 大肠杆菌素 K	以一种未知的作用中断膜上依赖能量的过程
大肠杆菌素 E2	降解 DNA
大肠杆菌素 E3	切割 16S　rRNA

Col 质粒分子质量只有几百万道尔顿，所以没有自身转移能力。研究得最清楚的 Col 质粒是 ColE 1，分子质量是 4.4×10^6Da。它的全序列为 6646bp。ColE 1 被广泛地用于 DNA 重组研究。

ColE 1 是一个具动员能力但无结合能力的质粒，它为质粒转移时需要由质粒编码的核酸酶提供了可靠的证据。ColE 1 的核酸酶由 *mob* 基因编码，酶专一作用的一段碱基顺序叫 bom，关键性实验如图 3-12。在图中，相容性质粒 F 和 ColE 1 共存于同一细菌细胞中，F 提供 ColE 1 缺少的结合功能，这样能够使 ColE 1 转移。图 3-12 中（a）表示 ColE 1 在 F^- 细胞中的状态，*mob* 基因转录产生 mob 产物使 bom 位点产生切口（这个切割位点叫 nic），ColE 1 超螺旋环状 DNA 变成切口环。因为 ColE 1 不能产生性纤毛形成接合对而不能转移。图 3-12（b）中细菌含有 F 和 ColE 1 两种质粒，F 合成纤毛和转移装置；ColE 1 可被转移。已经分离到了当 F 提供转移装置时 ColE 1 的突变体（mob^-）也不能被转移，遗传学分析表明 mob^- 突变是隐性的，这说明 *mob* 基因编码一种蛋白质，进而通过分离到的这种突变质粒发现它不是以松弛复合物形式存在。图 3-12 中（c）表明 F 不能帮助 mob^- 突变体转移，在这种情况中，虽然 F 能提供转移的装置，但 ColE 1 质粒 DNA 上不能产生切口，另一种 ColE 1 突变体表型为 bom^-，它是一种顺位显性突变，缺失了 bom 位点，mob 蛋白虽能产生但不能在质粒 DNA 上产生切口，因而也不能转移 [图 3-12（d）]。

(4) 土壤细菌质粒 Ti

在许多双子叶植物中发现的根茎擦伤瘤是由土壤细菌引起的。这种可引起肿瘤的活性存在于质粒 Ti 中，当植物被感染时，这种细菌进入植物细胞，在植物细胞中生长并裂解，将它们的 DNA 释放到这个细胞中。此后 Ti 质粒的一个含复制基因的小片段以一种还不清楚的机制整合到植物细胞染色体上，这个整合片段打破了控制细胞分裂的激素调节系统，使细胞变成肿瘤细胞。这种质粒目前在植物的育种中非常重要，特定的基因可以通过重组 DNA 技术插入到 Ti 质粒上，这

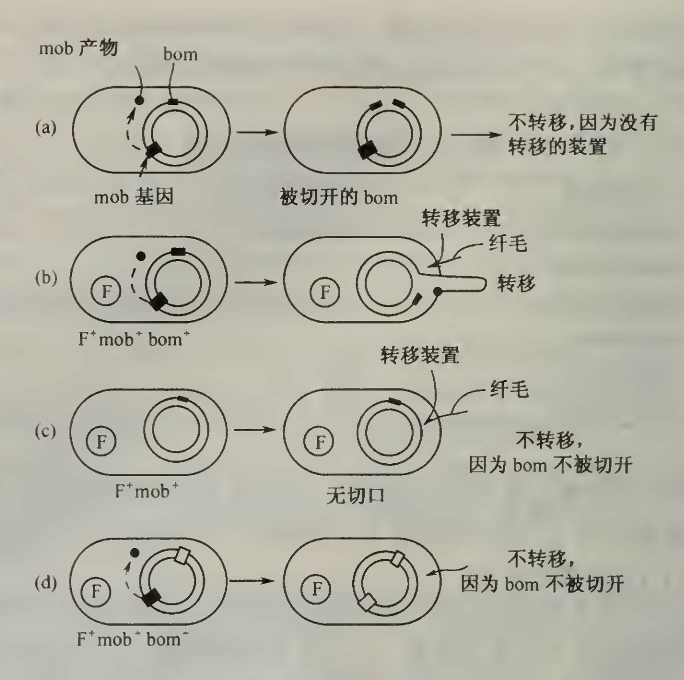


图 3-12　在一定条件下 F 因子能使质粒 ColE 1 变成可移的

些基因有时又可以整合到植物的染色体上，从而永久地改变这种植物的基因型和表型。可以用这种方法育出各种各样的人们需要的有经济价值的新植物品种。

(5) 真核生物中的质粒

已知单细胞真核生物酵母中存在质粒，最有趣的质粒之一就是 killer particle，它是分子质量为 15×10^6Da 左右的双链 RNA 分子，它是已知唯一不含 DNA 的质粒。这个质粒含有 10 个与复制有关的基因和几个合成杀伤物质的基因，这种杀伤物质类似于大肠杆菌素。

另一种研究得最清楚的酵母质粒是 DNA 长度为 2μ、分子质量为 4×10^6Da 的质粒，称为 2μ 质粒。这种质粒在每个细胞中的拷贝数是 60 个，并且都存在于酵母的细胞核中，与所有核内 DNA 一样，这种质粒 DNA 也和组蛋白结合在一起，质粒 DNA 显然不是整合到宿主 DNA 上的。2μ 质粒的碱基顺序包含两个反向重复序列，每个为 599bp；这两个反向重复顺序被两个单一顺序隔开，一个为 2346bp，另一个为 2774bp。在一个细胞里这种质粒有两类，这两类质粒由其中一个单一序列的方向来区分。两类质粒数目相等，由于反向重复顺序之间的频繁重组，两类可以互相改变，这种互变由质粒基因 *flp* 催化。互变的功能还不清楚，在实验室中 flp^- 质粒可在宿主酵母细胞中复制，显然没有缺陷。

2μ 质粒的复制具有与其他质粒明显不同的特征，它是高拷贝数质粒，其拷贝数可能也是由一个与大肠杆菌质粒 pSC 101 和 ColE 1 有关的系统来调控，但

是对于 2μ 质粒来说不存在这种情况，在细胞分裂的每个周期中每个质粒 DNA 分子只复制一次。它是如何调控的还不清楚。

大部分真核生物质粒像原核生物质粒一样是环状 DNA 分子，但从玉米、高粱和几种真菌中分离到的少数质粒是线状 DNA 分子，这样的分子有一个复制终止的问题，这也是所有线状 DNA 分子面临的问题。这种质粒的碱基序列测定显示出的独特的终止结构是回文、发夹和重复序列，很像酵母和四膜虫的端粒，也许这些质粒终止复制的方式类似于染色体中所用的方式。

二、转座因子

20 世纪 40 年代首先在玉米中发现了类似于现在的转座因子的解离激活因子（DS-AC）系统以后，直到 1967 年才在研究大肠杆菌的乳糖操纵子中第一次发现称为插入序列的转座因子。已在细菌中发现两类转座因子：复杂转座因子和简单转座因子。插入序列就是简单转座因子。

1. 插入顺序

插入顺序（insertion sequence，IS）可以在不同的复制子之间转移，以非正常重组的方式从一个位点插入到另一个位点，对新位点基因的结构与表达产生多种遗传效应，常常对插入点的基因的功能表达产生极性效应。

（1）插入顺序的发现

插入顺序最早是在 1966～1967 年研究大肠杆菌半乳糖操纵子和乳糖操纵子的自发极性突变时发现的。当时发现一些极性突变，在恢复突变方面与一般的点突变不一样。例如 *E. coil* 染色体上 *lac Z* 基因发生突变后，它后面的 *lacY*、*laca* 基因不能表达。

20 世纪 60 年代初期，在大肠杆菌中分离到一种乳糖操纵子的突变株。突变位点发生在 *lacZ* 基因上，它不但使细菌不能合成 *lacZ* 基因所编码的有活力的 β-半乳糖苷酶，而且也使 *lacY* 基因所编码的半乳糖苷透性酶和 *lac a* 基因所编码的半乳糖乙酰转移酶的合成量减少，即所谓的极性效应。另一些突变株的突变位点在 *lacY* 基因，这些突变株不能合成 β-半乳糖苷透性酶，而且半乳糖苷乙酰转移酶的合成量也减少了，但半乳糖苷酶的合成量却完全正常。

经研究表明是由于在突变位点上发生了一个碱基的突变，结果使转录后的 mRNA 产生了一个无义的密码子，蛋白翻译到此处时，肽链的延长终止，产生无酶活性的多肽片段，并且影响同一操纵子中后面的酶的合成。这种类型的突变称为无意义极性突变。

下面是证实插入顺序的实验过程。大肠杆菌的半乳糖操纵子是由 O 区、P 区及结构基因 *E*、*T*、*K* 组成的。*galE*、*galT* 和 *galK* 基因编码的酶催化使半乳

糖变为葡萄糖-1-磷酸的反应。

当 *galE* 基因发生突变时，半乳糖-1-磷酸与尿苷二磷酸半乳糖（UDP Gal）在细胞中积累，这两种糖磷酸酯对大肠杆菌有毒性，因此突变株不能在含有半乳糖的培养基上生长（对半乳糖敏感）。但将大量突变株涂布于含半乳糖的平皿时，也会长出对半乳糖产生抗性的少数菌落。研究证明，这不是由于回复突变，而是由于 *galT* 基因发生子突变，使该突变株不能合成 *galT* 基因所编码的尿苷酸转移酶，并且使 *galK* 基因编码的半乳糖激酶的合成量也减少，不能积累 UDP Gal，而半乳糖-1-磷酸的堆积量也降低到对细胞无毒性的水平，所以突变株能在含半乳糖的培养基上生长。这种对半乳糖有抗性的突变株也能回复成对半乳糖敏感的菌株，因此这种突变不是缺失突变。但是它们回复突变的频率不能像无意义极性突变那样可以通过碱基取代型或移码型诱变剂增加，所以它们也不是无意义突变和移码突变。除此之外的可能性是半乳糖操纵子内 DNA 的倒位、重复或由于半乳糖操纵子以外的 DNA 片段插入所引起的。要鉴定这三种可能性，需要测定半乳糖操纵子的 DNA 含量。因为倒位并不增加 DNA 含量；重复则增加 DNA 含量，但不出现新的 DNA 顺序；而操纵子以外的 DNA 的插入既增加半乳糖操纵子的 DNA 含量，也出现新的 DNA 顺序。

Jardan 在 1968 年和 Shapiro 在 1969 年分别比较了半乳糖操纵子与带极性突变的半乳糖操纵子的 DNA 的分子质量。他们制备 λdgal 和带极性突变的 λdgal 的 DNA，用密度梯度离心比较这两种 DNA 的密度后发现带极性突变的 λdgal DNA 比 λdgal DNA 密度大，说明极性突变的半乳糖操纵子的 DNA 分子质量的确是有所增加，这就排除了这种极性突变是由于倒位所引起的可能性。

Michaelis 在 1969 年用带极性突变的 λdgal 的 DNA 体外转录，将得到的 mRNA与亲株 λ dgal（无极性突变者）的 DNA 反复杂交，从中得到一段不能与亲株杂交但能与带极性突变的 λdgalDNA 杂交的 mRNA，这说明带极性突变的半乳糖操纵子 DNA 中增加的 DNA 片段不是半乳糖操纵子固有的一部分，即突变不是由于 DNA 重复所引起的，而是由于增加了半乳糖操纵子以外的 DNA 顺序所引起的。并且还有研究发现一种极性突变的 λdgal 所转录的 RNA 能与另一种极性突变的 λdgal DNA 杂交，表明此种顺序是专一的，即这些极性突变都是由于插入同一种 DNA 顺序所引起的。因此称这种极性突变为插入型极性突变。

Hirsch 等在 1972 年用异源双链法证明了插入型极性突变的插入顺序是 λdgal DNA 中的一段不能与其亲株 λdgal DNA 杂交的 DNA，它能在异源双链中形成一个单链环。根据异源双链所形成的单链环的大小可以测定出 IS 的大小，IS 的大小是 768～7500bp。IS 除带有转座必需的遗传信息外，不带任何其他遗传信息，如 IS1（图 3-13）。

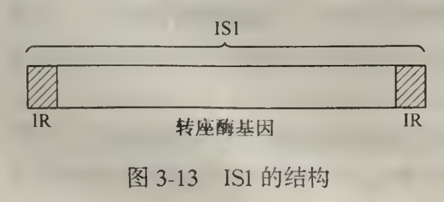


图 3-13　IS1 的结构

(2) 大肠杆菌染色体和 F 因子上的 IS

Saedler 等在 1973 年将带有极性突变操纵子（即含有 IS）的噬菌体 DNA 固定在滤膜上，同时用^3H标记细菌染色体，将制成的［^3H］DNA 与固定在滤膜上的含 IS 的噬菌体 DNA 杂交，测定膜上的^3H放射活性。用这种方法计算出 *E. coli* K12 染色体上有 3 个拷贝的 IS1，5 个拷贝的 IS2 和 5 个拷贝的 IS3。

Davidson 等在 1975 年用电镜研究 F 因子的异源双链，证实 F 因子的 0～17.6 区段存在 IS2 及 IS3，并且这个区段是 F 因子插入寄主染色体的热点（hotpoint），这些 IS 全部在重组区段中。图 3-14 表示 F 因子上的 IS 及其位置。

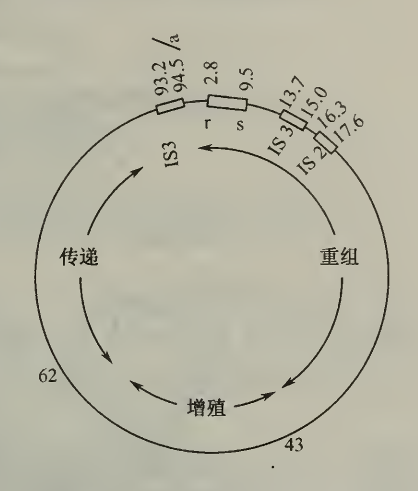


图 3-14　F 因子上的 IS 及其位置

数字以 kb 为单位。IS 集中在重组区。

(3) IS 在 Hfr 和 F 因子形成中的作用

Hfr 是 F 因子重组到细菌染色体上形成的。如图 3-15 所示，重组发生在 F 因子的 F_1、F_8 和染色体的 b_1、b_n 之间。发生传递时，F 因子的起点方向（图中箭头所示）决定移动方向。按图中所示的例子，F 因子插入后在染色体上标记的顺序是 b_3、b_2、b_1、F_1、F_2、F_3 …… F_7、F_8、b_n、b_{n-1}、b_{n-2}。F 因子如果按反方向插入则在染色体上标记的顺序完全颠倒过来，变成 b_3、b_2、b_1、F_8、F_7，……F_3、F_2、F_1、b_n、b_{n-1}、b_{n-2}。

如果发生与上述过程相反的反应，则 F 因子由 Hfr 重新回复到自主复制的状态。但在异常割离时，F 因子会失去一部分 DNA（图中失去 F_1），代之以染色体片段（b_{n-1}、b_n）而生成 F′因子。如果异常割离发生在染色体的两个部位上，那么 F 因子本身不丧失任何部分，生成取得染色体片段的 F′因子（图中为 b_n 和 b_1）。F 因子失去一部分的割离称为 I 型割离，F 因子没有丢失的割离称为 II 型割离。异源双链法研究证明 F′8gal，F′lac 等是 I 型割离；F′131ac，F′14ilv 是 II 型割离。

上面谈的是 Hfr 和 F′因子的形成。下面再来讨论 IS 在 Hfr 和 F′形成中所起的作用。Davidson 等将几种 F′因子与 F 因子并列排出，发现 F′因子中与寄主染色体的接合位点上都有 IS，这些因子中的 IS 位点有对应关系（图 3-16）。由图可见各种 F′因子与 *E. coli* 染色体的连接点上都有 IS 参加。

图 3-17 表示三种 F 因子怎样在 *proB*、*lac*、*proC*、*tsx*、*purE*、*lip* 等基因之间掺入。Hfrp804 是 F 因子上处于 93.2 位点的 IS3 和寄主染色体上的 $\alpha_3\beta_3$（IS3）发生重组而形成的；Hfr103 是 F 因子上处于 17.6 位点上的 IS2 和寄主染

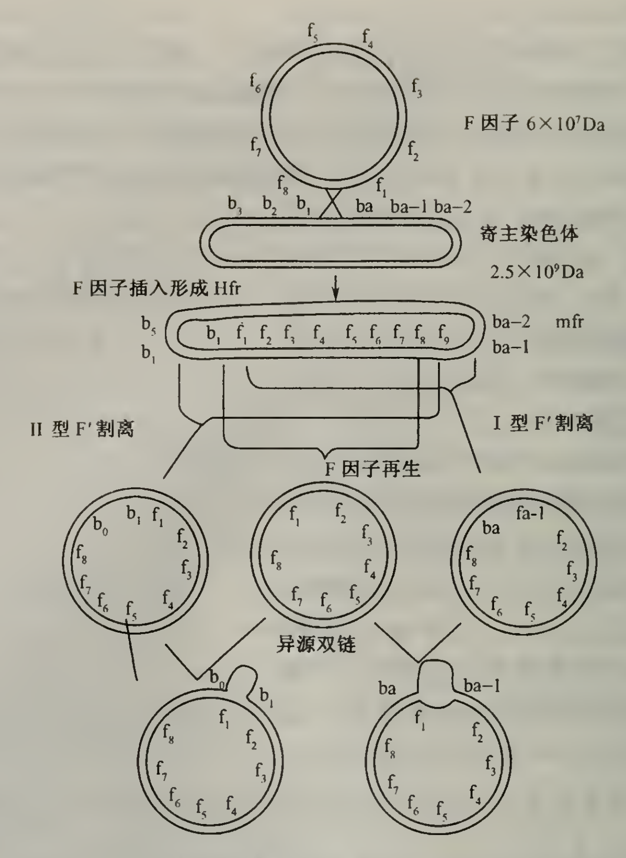


图 3-15　F 因子插入及割离模型、缺失一部分 F 因子 DNA 的是 I 型，完全不缺失 F 因子 DNA 的割离是 II 型

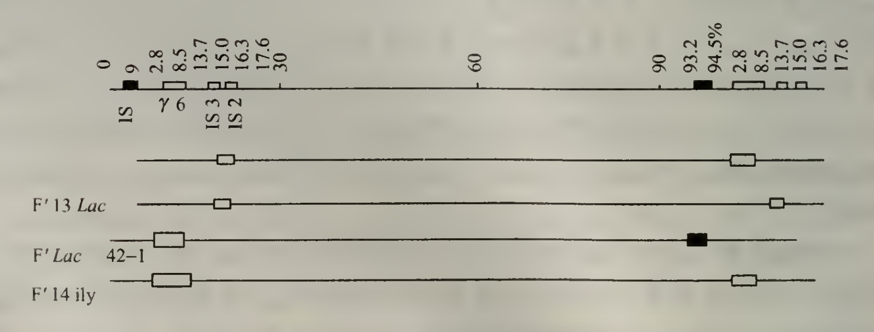


图 3-16　F 因子和染色体的连接点

色体上的 εδ（IS2）发生重组而形成的；Hfrp3 是 F 因子上处于 93.2 位点上的 IS4 和寄主染色体上的 $\alpha_4\beta_4$（IS4）发生重组而形成的。

由此可见 F 因子整合成为 Hfr 和割离形成 F′因子是以 IS 片段为媒介发生重

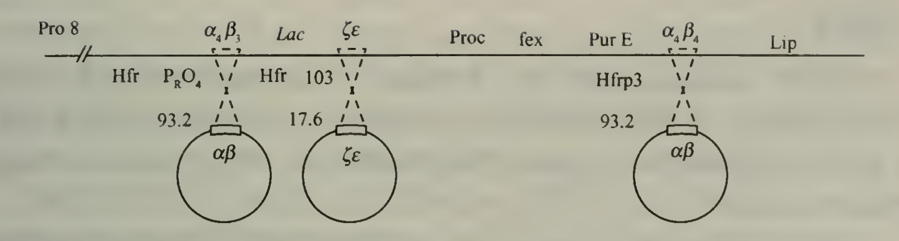


图 3-17 Hfrp804、Hfr103、Hfrp3 都是 F 因子和寄主染色体上的 IS 插入部位发生重组而形成的

组的。这就解释了为何 F 因子的 IS 都集中在重组区段而形成插入的热点。

(4) R 因子中的 IS

在抗药性因子 R 中发现存在 IS，例如 R6、R1 等均发现其中存在 IS。以上两种 R 因子在变成失去抗药性的 RTF（抗性转移因子）时，缺失了决定抗药性的一大段 DNA（抗性定子）。在抗性定子与 RTF 的结合点上发现存在 IS1。IS1 在它们的解离和再结合中起重要作用。

R1 是自然产生的复杂质粒，它含有转座子编码的抗氯霉素（Cm）和卡那霉素（Km）的基因，并且还含有完整的转座子 Tn4 携带的抗链霉素（Sm）、磺胺（Su）和氨苄青霉素（Ap）基因（图 3-18）。

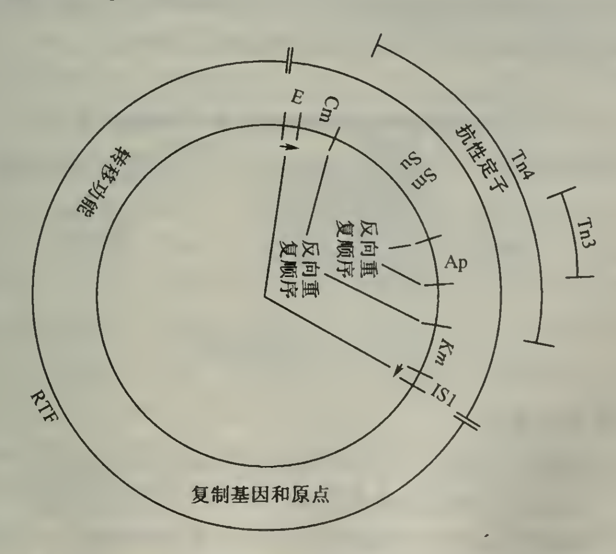


图 3-18 R1 是一个自然产生的 94kb 的复杂质粒，它抗 5 种抗菌素（Cm、Km、Sm、Su、Ap），由抗性定子和 RTF 两部分组成

2. 转座子

转座子（transposon，简称 Tn）又称易位子，由两个重复顺序夹着一个或多个结构基因组成，它们存在于 R 因子及其他质粒中。有的转座子的重复顺序就是 IS。转座子也和 IS 一样，能从一个位点转座到另一个位点，或从一个复制子转座到另一复制子。

在研究 R 因子的遗传特性时，发现一些具有不同抗性类型的 R 因子的抗性基因之间可以发生重组，甚至还分离到了 R 因子中的抗性基因整合到 P1 噬菌体基因组中的 P1CM 噬菌体。Datta 等 1971 年较详细地研究了这种现象，发现了 *Ap* 抗性基因可以转座，认为一个可以转座的 DNA 片段携带编码 β-内酰胺酶的结构基因，并称这一遗传结构为转座子 A。此后，一些实验室先后在不同的质粒中发现了几十种转座子。

(1) 转座子的结构及特征

转座子都比 IS 大，约 4500～200 000bp。将带有 Tn 的 DNA 变性，再进行链内复性，在电镜下可以看到一个有双链的茎和单链的环的 DNA 分子。研究证明单链环是 Tn 的结构基因部分，双链茎是两端的一对重复顺序，这些重复顺序多数是反向重复，因此在复性时形成双链的茎。用这个方法可测知一些转座子中结构基因的长度及其两端重复顺序的长度。所有转座子都有两个结构特征：①两端有 20～40 个核苷酸的反向重复顺序；②具有编码转座酶的基因，这种酶催化转座子插入新的位置。以 Tn3 为例，其结构如图 3-19 所示。

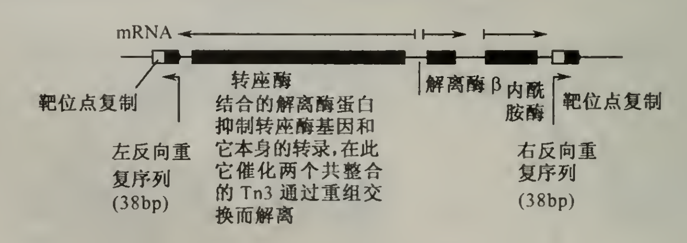


图 3-19　Tn3 的结构

(2) 转座频率的调节

细胞内的转座酶浓度很低，极少引起转座。每个细胞每世代产生的转座酶分子还不到一个，引起自发转座的频率只有 10^{-7}次。但是如果由含 Tn10 转座酶基因的工程质粒来提供更多的转座酶的话，Tn10 的转座频率可以提高 1000 倍以上。所以，很可能是转座酶的水平调控转座子转座的频率。

最近发现了决定转座酶表达与功能的一种重要因素。Tn10 和另外一些转座子的转座酶的启动子 DNA 含有 GATC 序列，在大肠杆菌中这一序列的 A 是甲基化的。因为 GATC 的互补链也是 GATC（都从 5′→3′读），此处 DNA 的两条链都

甲基化，在此情况下，启动子便被相对地灭活。但在DNA复制时，新合成的GATC中的A不立即甲基化（可能有5min或更长），这种一条链上甲基化另一条链上没有甲基化的启动子比上述两条链都甲基化的启动子活跃，于是在DNA刚复制后出现少数的转座酶。与此类似，在转座酶的作用位点上也含有GATC，新复制的DNA是转座酶的很好的底物，因此Tn10在DNA刚复制后易发生转移。目前尚不知道是否细胞其他条件影响DNA甲基化而改变转座的频率，也不清楚转座过程是否还受别的方式调节。

一些转座子还编码第二种酶——解离酶，它催化转座过程的第二阶段（图3-19），对于Tn3来说，解离酶蛋白还具有第二种独立的活性，它是自身基因和转座酶基因的抑制物，用于维持转座酶正常的低表达水平。它是通过结合在转座子DNA上的这两个基因之间的一个结合位点来抑制这两个基因的表达。

(3) 转座子如何移动

细菌转座子如Tn3的DNA及与其相邻接的DNA的序列分析为转座子转移机制提供了强有力的证据：①我们知道转座子携带它在插入时的所有序列准确转移，并不携带邻接的任何DNA。这只有在转座酶能识别转座子两端的情况下才能发生，因此两端的序列应是相同的，因为它们是同一种酶的结合位点。②在目标DNA的转座子插入位置上准确复制3～12碱基的目标DNA，造成一个插入位点，其中一份留在转座子的两端。因此在转座过程中必须发生DNA复制，创造出双份目标DNA，造成插入位点。这一过程与λ噬菌体整合过程中的位点特异性重组是明显不同的。③大多数转座子在移动到新的基因组的任一区域时，它们的移动并非完全是随机的，而是偏向于入侵一定的DNA序列。有些转座子攻击特殊的4个或6个碱基对的对称序列，如同限制性内切核酸酶切割DNA那样。

通过对一些能将DNA链打开并重新连接的酶如λ噬菌体的整合酶和DNA拓扑异构酶的研究，使人联想到一种解释转座酶如何工作的模型（图3-20）。转座酶结合于转座子的两端，同时也与目标DNA上的相应的序列结合，然后像限制酶那样交错切开目标DNA的两条链，同时也在转座子的两端的一条链上切开，但两端切的并不是同一条链，转座子的两条链的自由末端分别和目的DNA的两个插入点的末端连接在一起，通过转座子两条链的桥梁作用，新老两个目标DNA分子就连接在一起。由于生长的具体条件不同，转座子有两种选择：一种是简单转座，在转座子与原目标DNA连接的末端再切出两个切口（一条链上一个），转座子完全转座到新的目标DNA分子上并且在原来的宿主目标DNA处留下一致死性的缺口，通过DNA聚合酶的作用插进几个核苷酸完成目标位点的倍增［图3-20（c）和图3-20（d）］。因此简单转座是一个全保留的过程。第二种可能是复制性的转座过程，在此过程中，转座子是倍增的。一个拷贝转移，另一个拷贝留在原来的宿主DNA上，此过程中宿主DNA没有被弄断，转座子DNA没有被第二次切开，而是留下来由DNA聚合酶复制。DNA聚合酶从第一次切开处

开始复制，产生了叫共整合（cointergrate）的结构，共整合结构的存在是这种转座过程的有力证据。解离酶以一种类似λ噬菌体整合的反应方式催化两个转座子拷贝之间的位点特异性重组，产生两个各含一个转座子拷贝的分离开的 DNA 分子。

这些步骤可以清楚地解释转座子的其他活动，例如，假使没有解离酶，那么终产物就是两个分子连在一起的结构，或两个连在一起的分子错误地分开成两个独立的重排的 DNA 分子。图 3-20 左边的模式表明，转座到同一分子的另一位置的过程中会造成倒位和缺失。

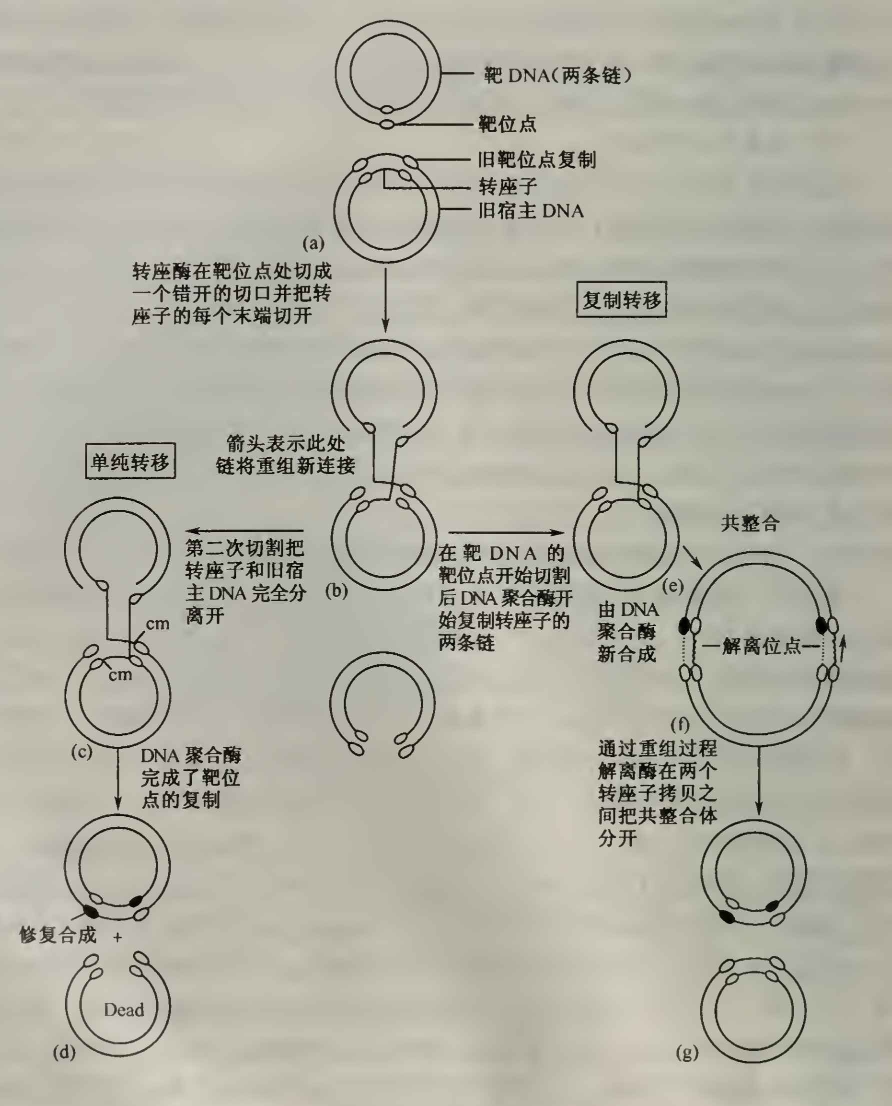


图 3-20　转座模型

要彻底弄清转座的机制，需要在体外对每一个步骤进行研究。由于细菌噬菌体 μ（它既是噬菌体又是一种转座子）的转座过程在无细胞体系中也可以发生，我们希望这一发现能使我们进一步精确地了解转座子的转座过程及其到底是如何移动的。

最近的实验表明，一些较高等的生物的转座子，尤其是反转病毒，转移方式与细菌的转座子如 Tn3 转移方式差别很大。低等真核生物酵母菌的转座因子 Ty 是先转录成 RNA，再反转录成 DNA，然后这种 DNA 以一种尚不清楚的机制重组到宿主染色体的新的位置中去。酵母的 Ty-I 转座子也称为 retrotransposon。

(4) 转座因子的特征和遗传学效应

转座因子的共同现象以及它们带来的遗传学效应有以下几个方面：

①以 10^{-8}～10^{-3}频率转座，引起插入突变。各种 IS、Tn 和 Mu 噬菌体都可以引起插入突变，如果插入的位置是一个操纵子的前端的基因，那么可能造成极性突变。一般的碱基置换突变没有极性效应，只有终止密码突变和移码突变才有极性效应；通过 IS1 和 IS2 的碱基序列分析发现它们确实存在无义密码子。IS2 以任何方向插入都有极性效应，可是 IS3 只有以一个方向插入时才有极性效应，相反地也有关于由于 IS2 插入而出现新的启动子的报道。

②插入位置上出现新的基因。如果转座子上带有抗药性基因，那么它一方面造成插入突变，另一方面在这一位置上出现一个新的抗药性基因。

③造成插入位置上出现受体 DNA 的少数核苷酸对的重复。例如 IS1 插入基因 *lacI* 后造成 9 个核苷酸对的重复出现。

ACGAT　GTCGCAGAG　TATGC

lacI

TGCTA　CAGCGTCTC　ATACG



IS1

ACGAT　GTCGCAGAC　GTCGCAGAG　TATGC

IS1

TGCTA　CAGCGTCTC　CAGCGTCTC　ATACG

不同的转座因子插入后造成重复碱基对数不等，例如 IS1、Tn10 等造成 9 个碱基对的重复，IS2、Tn3、Mu、$\gamma\delta$ 等造成 5 个碱基对的重复，IS3 造成 3 个或 4 个碱基对的重复，IS4 造成 11 个碱基对的重复。

④转座后原来位置上保持原有的转座子。

⑤转座因子插入染色体后引起其旁边的染色体畸变。最初在 IS1 中发现由于 IS1 的存在促使它的旁边发生缺失。缺失发生的频率高于自发缺失突变频率的 100～1000 倍，以后在 IS2、Tn3 等转座子中都发现有这一现象。

转座子插入染色体后引起染色体畸变的原因可能是由于转座子首先以相反或

相同方向转座到邻近位置上，然后通过某种方式的重组产生倒位或缺失（图3-21）。

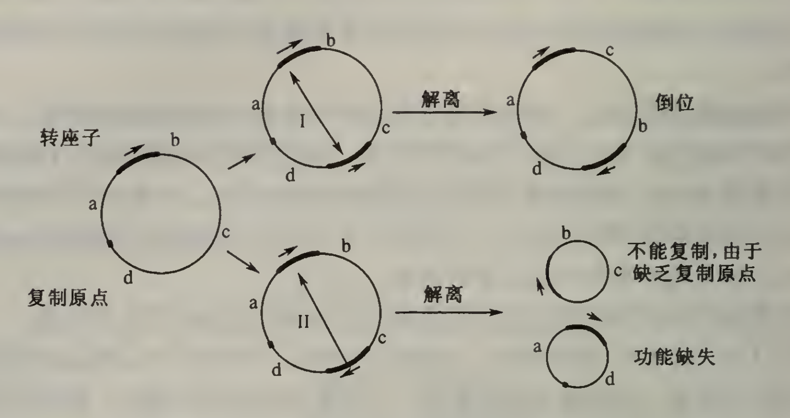


图 3-21　转座因子引起缺失和倒位模型

⑥切离。转座子可以从原来位置上消失，这一过程称为切离。切离可以是准确的，准确切离的结果使发生了插入突变的基因恢复突变，如果这个转座因子是一个带有抗药性基因的转座子，那么抗药性同时消失。不准确切割的结果使发生了插入突变的基因不能恢复突变，转座子本身所带的遗传标志消失，这是不准确切离引起染色体畸变的结果。这些现象均已在插入 Tn10 的 hisG 突变株的恢复突变中得到了证明。

⑦核苷酸顺序。如前所述，各种已知的 IS 的长度在 768bp（如 IS1）到 7500bp（如 Y6）之间，它们两端都有 18～41bp 反向重复序列。已知转座子的长度在 2088bp（如 Tn1681）到 20 500bp（如 Tn4）之间，它们的两端有反向重复序列，某些转座子的两端的重复序列是已知的 IS，如 Tn9 和 Tn204 的两端是 IS1，也有一些转座子的重复序列不是已知的 IS。

IS1 的全部核苷酸顺序已经在 1978 年测定，全长为 768bp，两端是 35bp 的反向重复序列。

由于 IS1 的两端是反向顺序，所以不论 Tn 两端的 IS1 是反向还是顺向的，它们都有反向序列，似乎两端的反向序列使它们具有转座子功能，而且许多真核生物的转座因子也有同样的结构特性。不过也有例外，如酵母菌的转座子 Ty1 的两端没有反向重复顺序，Mu 两端没有重复顺序。

（5）转座因子在遗传学中的应用

转座因子除了本身在遗传学中的意义以外，在许多情况下是遗传研究中的有用工具。N. Kleckner 等在 1977 年总结了这方面的应用，下面列举几个主要方面：①用于难以筛选的基因的转移。②作为基因定位的标记。③筛选插入突变。④构建特殊菌株。⑤克隆难以进行表型鉴定的基因。⑥F 因子处在特定位置上的

Hfr 菌株的构建。⑦带有特定基因的转导噬菌体的构建。⑧特定区段缺失的菌株的构建。

三、病毒及其与质粒、转座因子之间的关系

最早将病毒描述为仅在细胞中扩增并且能通过最小的细菌滤器的致病因子。所有病毒都含核酸，有些病毒核酸是 DNA，有些是 RNA（图 3-22）。这些病毒核酸包装在蛋白质外壳中，有些病毒有包膜，几种病毒见图 3-23。病毒基因组的结构和复制模式在病毒中存在很大的差异。病毒在宿主细胞中扩增最终摧毁宿主细胞，病毒感染细胞后一般使细胞裂解释放有感染性的病毒粒子。有些病毒感染宿主后将其基因组整合到宿主染色体上，成为宿主染色体的一部分，随宿主染色体复制而复制，这种情况中的病毒称为原病毒。

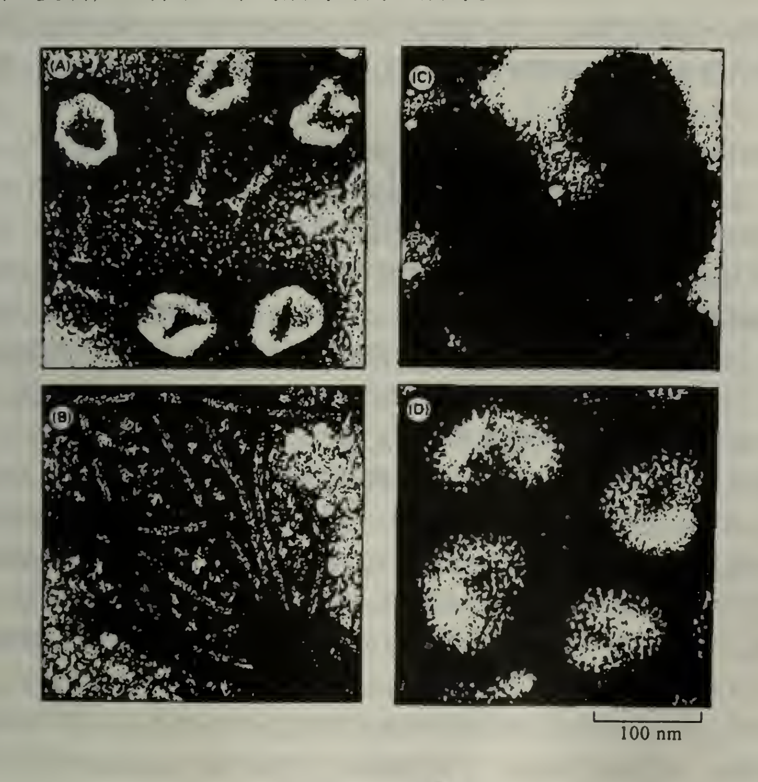


图 3-22　病毒囊膜

由于许多病毒可以以整合方式进入染色体，也可以割离后而离开染色体，所以一个大型基因组上可能含有多个不同的原病毒，多数基因组也可能含有多种不能形成病毒粒子和不能离开细胞的可移动的 DNA 序列，例如真核基因组中存

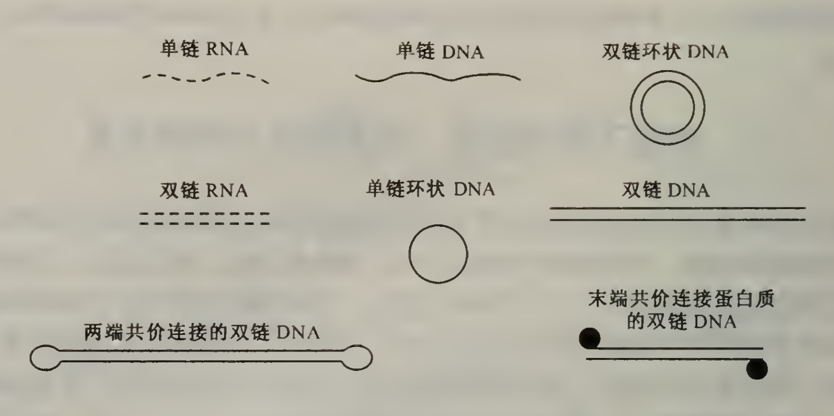


图 3-23 病毒基因组的几种类型

在的长短不一的重复顺序，大小在几百到几千个碱基对的转移因子在一个细胞里常以多拷贝存在，这些转移因子可在它自身编码的酶作用下转移位置，但转座频率很低，大约在 10^{-5}～10^{-4}。

转座机制多样。转座因子的一个大家族利用反转病毒生活周期中某一部分机制转座，这种转座因子叫反转座子（retrotransposon）。酵母中的 Ty-1 因子研究得最为清楚，转座的第一步是完整的转座因子的转录，产生一个 5000 多碱基的 RNA，该 RNA 编码反向转录酶，通过反向转录，合成双链 DNA（dsDNA），这和反转病毒感染后的早期步骤一样，该 dsDNA 在整合酶作用下，整合到染色体上的随机位点。虽然它们和反转病毒的相似性引人注目，但也有与反转病毒不同的地方，Ty-1 没有蛋白质外壳，它只能在一个细胞内移动并产生子代。因此认为一些可转移因子和反转病毒关系密切。

大多数病毒可能由质粒进化而来，最大的病毒也依赖于宿主细胞进行生物合成，因而细胞必须在病毒之前进化。第一个病毒的前身可能是一个小的核酸片段，逐步发展成具有独立于宿主细胞染色体复制的能力，这种能自主复制的因子称为质粒，能在细胞染色体外自主复制。质粒有 DNA 和 RNA 两种形式，这一点像病毒，它们含有一段特别的核酸序列作为它们的复制原点。与病毒不同的是，它们不能制造衣壳蛋白，所以它们不能像病毒那样在细胞间移动。

第一个 RNA 质粒与在一些植物细胞中发现的类病毒类似，这些仅仅 300～400 核苷酸大小的环状 RNA，尽管不编码任何蛋白质，但可复制，没有衣壳，只能通过受体和供体的受伤的细胞表面之间传递，因受伤处无细胞膜保护。这些可自主复制的因子可能从宿主细胞获得一些核苷酸序列，可以利于它们的复制，也包括编码衣壳蛋白的序列。目前发现的有些质粒确实相当复杂，可以编码控制它们复制的蛋白质和 RNA 分子，也编码控制它们进入子代细胞的蛋白质。已知最大的质粒是大于 100 000bp 的双链环状 DNA。

当一个 RNA 质粒获得编码衣壳蛋白基因后，就可能出现第一个病毒。但一个衣壳只能把有限的核酸包在里面，因此，一个病毒所含有的基因的数目是有限的。病毒被迫最有效地利用它们有限的基因组，一些小病毒发展重叠基因，在同一 DNA 序列内利用不同的阅读框编码两种以上的蛋白。

当病毒能在不同种宿主之间转移它们的核酸后，它们便在它们感染的宿主生物的进化中起了重要作用，许多重组经常是在它们的不同种的宿主基因组之间进行，通过这种方式，它们随机地得到宿主的一小段染色体，并把它带到不同的细胞或机体中去，有些整合的病毒 DNA 成为大多数宿主基因组的正常部分，例如 λ 噬菌的整合，另外，在脊椎动物基因组中发现了大量拷贝的内源性逆转病毒。整合的原病毒可不再产生完整的病毒，但仍能产生它所编码的蛋白质，有些蛋白对宿主是有用的，因此，像有性生殖一样，病毒通过促进基因库的混合而加速进化。

通过病毒使 DNA 序列在不同的宿主细胞基因组之间转移的过程叫做 DNA 转导（DNA transduction），有几种能高效转导 DNA 的病毒用于基因在两种细胞间的转移。

第五节　基因和基因组

基因一般是指表达一种蛋白质或功能 RNA 的遗传物质基本单位。

原核生物的基因组就是它的整个染色体。对于一般的二倍体高等生物来说，能维持配子或配子体正常功能的最低数目的一套染色体构成一个基因组。

在 20 世纪 50 年代以前人们认为一般情况下每一基因组的 DNA 的量是固定的，它包括数目固定、位置固定、功能固定的一系列基因，而且这些基因的排列是随机的。1961 年大肠杆菌乳糖操纵子的发现说明大肠杆菌的染色体不是一个随机排列的基因集合体，它由许多操纵子以及一些单独的基因组成。

关于基因位置的固定性问题，在 20 世纪 40～50 年代研究玉米的控制因子时已经指出某些遗传因子可以转移位置，可是这一现象当时并没有受到重视。20 世纪 60 年代末在大肠杆菌中发现了可以转移位置的插入序列，于是人们开始认识到基因组中的某些成分位置的不固定是一个普遍现象。

插入序列和转座子从一个位置转移到另一位置时，原来位置上的这些结构依然存在，这就是说实际上转移的是一个复制品，因此可以想像一个基因组的 DNA 的量由于转座子的存在可以继续发生变化。事实上用 Southern blot 和分子杂交法测得同一种生物的不同个体的基因组上同一插入序列或转座子的数目不相等，于是又使人们认识到一种生物的基因组的大小或者基因数目的多少并不是绝对固定的。

在 1922 年已经发现沙门氏菌的鞭毛相的转变现象。近年来发现相转变和基

因组上一小段 DNA 的位置颠倒有关，这一事实除了说明细菌的基因组的结构并不是绝对不变的以外，还说明随着结构的改变基因的功能也在发生变化。

近年来随着重组 DNA 技术的发展，基因组结构的研究也迅速发展起来。这些研究成果促进了人们对基因组结构的深入了解。目前人类基因组计划、全序列业已完成，正在进行后基因组研究工作。

原核生物的基因组的特点是：①小，一般具有单一 DNA 复制起点；②单个染色体，一般呈环状；③染色体 DNA 或 RNA 并不与蛋白质形成固定的结合物；④少量重复序列；⑤功能上密切相关的基因构成操纵子或高度集中，并且常转录成为多基因 mRNA。

愈是简单的生物，它的已知基因数愈是接近于根据它的 DNA 分子质量所估计的基因数，它的 DNA 的含量测定也愈准确。对多种噬菌体和细菌基因组的研究表明它们的基因组中除了控制区以外没有不编码蛋白质的 DNA。

真核生物基因数不确切。粗糙脉孢菌、果蝇等基因组中编码蛋白质的 DNA 估计为 5%左右。人的基因数姑且假定是 30 000，那么人基因组中编码蛋白质的 DNA 大约为 0.5%，即使假定基因数是 300 000，编码蛋白质的 DNA 也不过 5%。

原核生物中很少有不编码蛋白质的 DNA 序列，真核生物中进化位置愈高的生物不编码蛋白质的 DNA 序列愈多。

原核生物的基因组的另一特点是功能相关的基因高度集中而且愈简单的生物似乎愈集中。

λ 噬菌体的 7 个头部基因 $A \sim F$ 占有相互邻接的位置，调控基因 *N*、*CI*、*CII*、*CIII*、*Cro* 也集中在一个区域。此外，结构基因与它们所编码的蛋白质所作用的部位也在邻接的部位，例如整合基因 *int* 和切离基因 *xis* 处在附着位点 att 的旁边；复制基因 *O* 和 *P* 处在复制起始点旁边。

细菌的基因集中程度总的来说不如 λ 噬菌体，但与真核生物相比则比较集中，在作用上密切相关的基因都组成操纵子，真核生物中虽然也有较多作用上密切相关的基因集中在一起的情况，可是关于操纵子的报道还绝无仅有。

原核生物和真核生物的基因组的另一区别是重复序列的数量。原核生物的基因组中有较少的重复序列，而真核生物基因组中则存在大量的重复序列。

第四章　DNA复制

无论是只含一个染色体的细胞还是具有多个染色体的细胞，在细胞分裂之前，其完整基因组DNA必须预先复制。复制如何与细胞分裂周期相关?

1) DNA复制的起始保证了细胞进一步的分裂，从这个基本点出发，一个细胞产生的子代的数目是由一系列控制DNA复制是否起始的调节因子决定的。

2) 如果复制起始，在复制过程完成之前不允许细胞分裂的发生。DNA复制的全部完成可能给细胞分裂提供了一个开关（trigger)，之后这个双倍基因组分配到每个子代细胞中（真核中的有丝分裂)，分离的单元是染色体。

在原核细胞中，复制的起始与细菌染色体DNA上一特别位点相关，分裂过程通过一个隔膜的发育完成。在真核细胞中，DNA复制的起始和S期的开始相一致，而且这个过程牵涉到许多个别的起始事件，通过有丝分裂中细胞的重新组织（reorganization)，建立细胞的分裂活动。

第一节　DNA的复制起始区

也许每个细胞必须做出最重要的决定：是否合成和在什么时候合成自己的DNA。类似于RNA的合成和其他生物学过程，复制的起始是被控制的。如果在体内DNA的复制总是从一特定位点起始，对复制的起点进行最有效地控制。电镜研究显示，动物病毒DNA复制泡的中心总是几乎位于同一位点，同样的结果也在细菌、植物病毒和各种质粒的环状DNA中获得。更详细的分子生物学研究指出，实际上DNA的复制起始于复制泡中心位置附近的特定的碱基序列，称为复制原点。

为了了解复制原点的结构特点和一般性质，在此将讨论三类复制原点，大肠杆菌（*E.coli*）OriC、酵母的自主复制顺序（yeast autonomously replicating sequences ARS）和猴病毒SV40的复制原点。

大肠杆菌的复制原点OriC是存在于大肠杆菌DNA复制起始位点的一段大约240bp的DNA片段，含有OriC的质粒或任何其他环状DNA分子均可在大肠杆菌中自主且有控制的复制。将OriC与其他五种不同的细菌其中包括亲缘关系远的*Vibrio harreyi*（一种海生细菌）相比较，发现它们都含有9bp的重复序列和13bp的A-T丰富序列，归类为9聚体（dnaA box）和13聚体（图4-1)，这些是起始复制的DNA蛋白识别结合盒（box)。另外，*E.coli*染色体上含有一段与OriC相邻的A-T相对丰富的DNA片段，这些DNA容易熔链而使复制机器在此

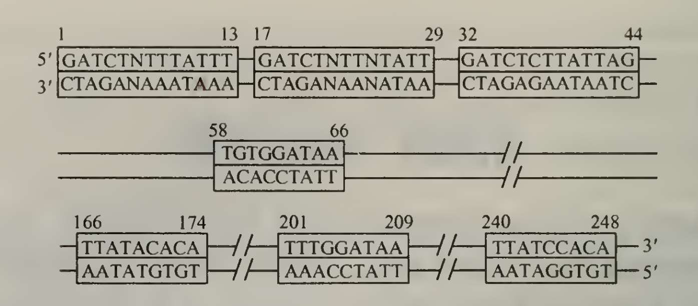


图 4-1　根据对六种细菌基因组的分析，细菌复制原点存在很短的一致性序列

得以运行。因此，这段 DNA 对复制的起始很重要。

E.coli DNA 复制起始以后，两个子代双链 DNA 上的复制原点与细菌细胞质膜上两个不同位点上的特异性蛋白质相连接，这种连接保证了细胞分裂时新合成的子代双链 DNA 分子分配到两个子代细胞中（图 4-2）。

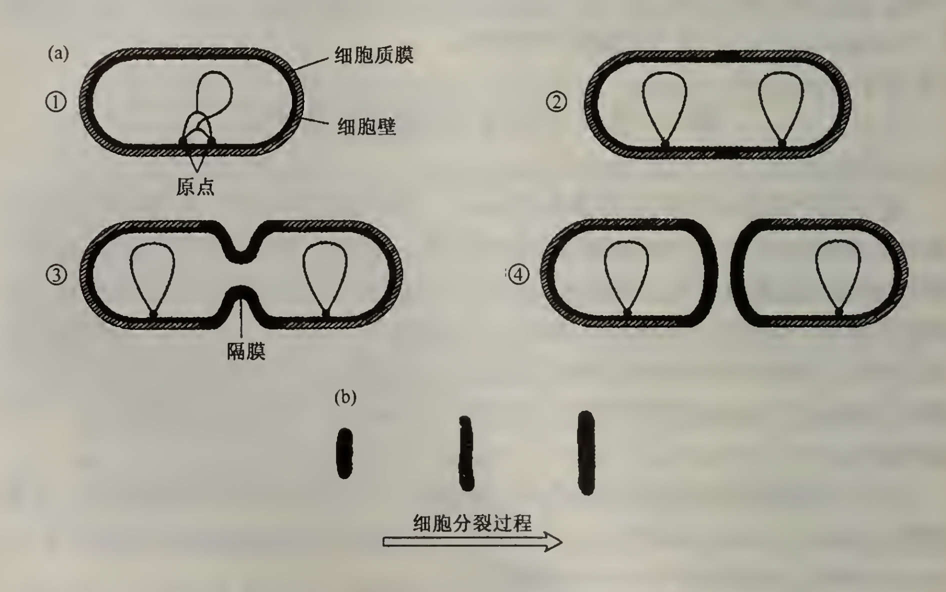


图 4-2　原核中的 DNA 复制和细胞分裂
(a) DNA 复制；(b) 细胞分裂。

酵母的自主复制序列。酵母的自主复制序列（ARS）就是酵母 DNA 的复制原点。与所有的真核染色体一样，每个酵母染色体有许多复制原点。分子克隆实验证明在酵母 *S.cerevisiae* 的 17 条染色体上存在 400 个复制原点；特征经详细鉴定的有 12 个以上。酵母的复制原点 ARS 对于 DNA 分子在酵母中的复制是必

需的。用精细突变分析方法研究一个称作 ARSI 的大约 180bp 的 ARS 的结果显示，仅有一个必需的元件，长度仅仅为 15bp，定名为元件 A，从 114 位到 128 位。还有三个其他小元件——B1、B2 和 B3 元件，它们可以提高 ARSI 发挥功能时的效率。比较了许多起 ARS 作用的、功能必需的不同 DNA 片段，发现它们都含以下 11bp 的保守的识别顺序：

5′A/T-T-T-T-A-T-A/G-T-T-T-A/T3′

ARSI 中的元件 A 在这 11bp 保守顺序中有 10 位是相同的，元件 B2 在 11bp 中有 9 位相同。

DNase 足迹法研究表明，一个称为 ORC（origin recognition complex）中有六种不同蛋白质复合物特异地与 ARSI 中的元件 A 结合，并且也特异地与其他 ARS 结合。在整个细胞周期中，该复合物一直维持与一种 ARS 结合在一起，并在 DNA 复制期间与其他复制相关蛋白联合，这是打开 DNA 复制起始开关的一个明显事件。所有真核细胞中的这些蛋白质都是同源的，这证明了它们在 DNA 复制中的重要性。

无论在哺乳动物细胞内还是在体外系统中，SV40 的复制原点都足以启动 DNA 的复制。突变研究证明 SV40 复制原点中有三个片段是启动活性所必须的，用带有 SV40 复制原点的质粒 DNA 和哺乳动物细胞内蛋白质可以研究 DNA 复制起始的分子机制。

虽然 *E.coli*、酵母和 SV40 的复制原点的特异的核苷酸顺序很不相同，但它们有几个共同的特征。第一，复制原点是含有多个短重复顺序的独特的 DNA 片段；第二，这些短重复顺序可被多聚原点结合蛋白所识别，这些蛋白质对 DNA 聚合酶与其他复制酶在复制起始位点上的装配有关键作用；第三，复制原点区通常含有 A-T 丰富区，这使得双链 DNA 解旋更容易。原点结合蛋白通过指导复制机器在染色体的特殊位点上的装配来控制复制的起始。

第二节　DNA 复制的相关蛋白质

在 DNA 复制中，酶促合成是个复杂的过程，因为这个过程需要高度准确的复制，并且在复制过程中亲代链必须自然分离，完成这个复杂过程的步骤太多，不可能由一个酶来完成。事实上，现已知道约 20 多种蛋白质参与复制。表 4-1 是与大肠杆菌复制有关的蛋白质。

一、DNA 聚合酶

DNA 聚合酶是以脱氧核苷三磷酸为底物催化合成 DNA 的一类酶。在不同的

表 4-1　与大肠杆菌复制有关的蛋白质

蛋白质名称*	分子质量/10^3Da	亚基数目	功　能	每个菌中的分子数	蛋白质含量(以细胞湿重计)/(mg/kg)
SSB	74	4	与单链结合	300	20
i 蛋白	80	4	预引发	50	0.5
n 蛋白	25	1	预引发	30	
n′蛋白	75	1	部位识别, ATPase	70	0.3
n″蛋白	11	1	预引发	–	
dnaC	29	1	预引发	100	
dnaB	300	6	移动启动子, ATPase	20	0.3
引发酶	60	1	引发合成	50	0.2
pol III 全酶					
α	140	1		—	
β	25	1		—	
θ	10	2		—	
β	40	1	链的延长	300	0.5
γ	52	2		20	
δ	32	1			
τ	83	1			
pol I	109	1	填补缺口和切去引物	300	10
连接酶	74	1	连接	300	10
拓扑异构酶 II	400	4		–	
α	210	2	引进超螺旋	250	
β	190	2		25	
rep 蛋白	65	1	解开双链	50	0.6
解链酶 II	75	1		5000	
dnaA	48		复制起始	200	

* SSB, 单链结合蛋白; pol, 聚合酶; rep 蛋白是一种大肠杆菌解链酶。

生物包括细菌、植物、动物中都发现了这种酶。所有 DNA 聚合酶的作用方式基本相似。它们催化脱氧核苷三磷酸加到复制中的 DNA 链的 3′羟基末端，合成方向从 5′→3′，由模板决定加上何种脱氧核苷酸。DNA 聚合酶催化所需要的条件是：四种脱氧核苷酸、镁离子、模板 DNA 和 RNA 引物。还没有发现一种不需要引物就能重新合成 DNA 的 DNA 聚合酶。

1. 大肠杆菌 DNA 聚合酶 I

大肠杆菌 DNA 聚合酶 I 是 1956 年 Kornberg 及其合作者发现的第一个 DNA 聚合酶，故称为 Kornberg 酶。在各种 DNA 聚合酶中，这种酶研究得最透彻。它的一些特点代表了 DNA 聚合酶的基本特点。

纯的 DNA 聚合酶 I 由分子质量为 109 000Da 的一条肽链构成，每个大肠杆菌细胞中大约含 400 个酶分子。用枯草杆菌蛋白酶处理该酶，可获得两个片段，大片段分子质量 76 000Da，保留聚合酶活性和 3′→5′外切核酸酶活性，大片段又叫做 Klenow 片段，广泛使用于 DNA 序列分析中。小片段分子质量 34 000Da，具有5′→3′外切核酸酶活性。

DNA 聚合酶有多种活性。

（1）聚合活性

在模板指导下，以脱氧核苷三磷酸为底物在引物的 3′-OH 末端加上脱氧核苷酸。每个酶分子每分钟添加 1000 个单核苷酸。如果 DNA 的合成是由聚合酶催化的，那么需要下面的条件：①模板，②必须有一段 3′-OH 末端的引物，并且此引物必须与模板正确形成氢键，③合成方向从 5′→3′进行。如图 4-3 和图4-4 所示。

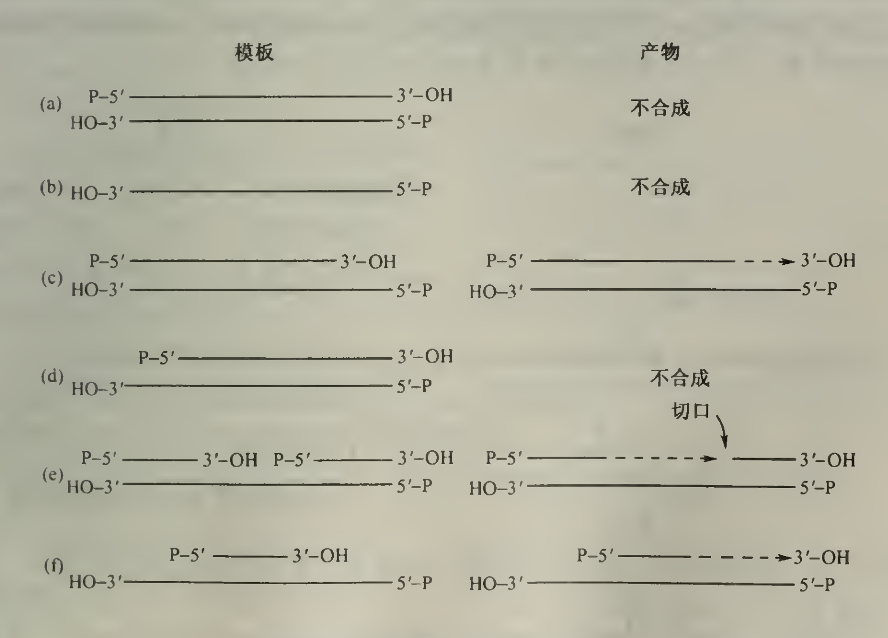


图 4-3　不同的模板聚合反应的效果

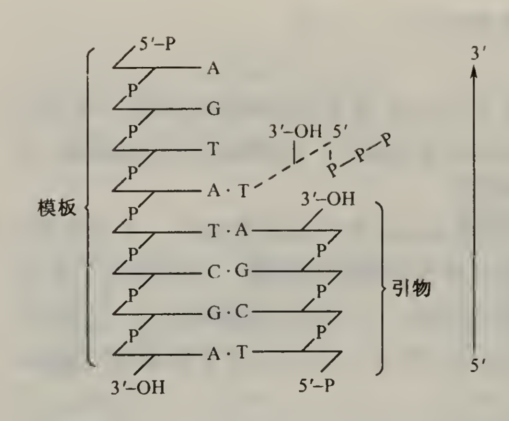


图 4-4　DNA 复制中模板引物和合成方向

(2) 3′→5′外切核酸酶活性

在没有脱氧核苷三磷酸底物时，能从 3′-OH 开始以 3′→5′方向水解 DNA，产生 5′-单核苷酸。实际上 DNA 复制时，加上去的脱氧核苷酸不一定每次都正确，有许多错误的机会，有时甚至加上一个与模板不配对的核苷酸，当 3′-OH 末端的碱基与模板不配对时聚合酶就无聚合活性，而它具有的 3′→5′外切核酸酶活性可以切除这个不配对的核苷酸，然后再恢复聚合活性，如图 4-5 所示。所以聚合酶的 3′→5′外切核酸酶活性也叫修补活性。

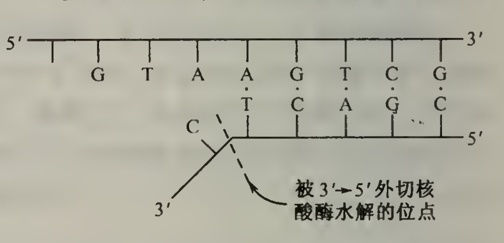


图 4-5　DNA 聚合酶 I 的 3′→5′外切核酸酶活性

(3) 5′→3′外切核酸酶活性

从图 4-6 中可以看到 DNA 聚合酶 I 具有 5′→3′外切核酸酶活性，其特点是：①只能在 5′-P 末端一个接一个地切除核苷酸；②能连续地切除多个核苷酸；③只切除配对的 5′-P

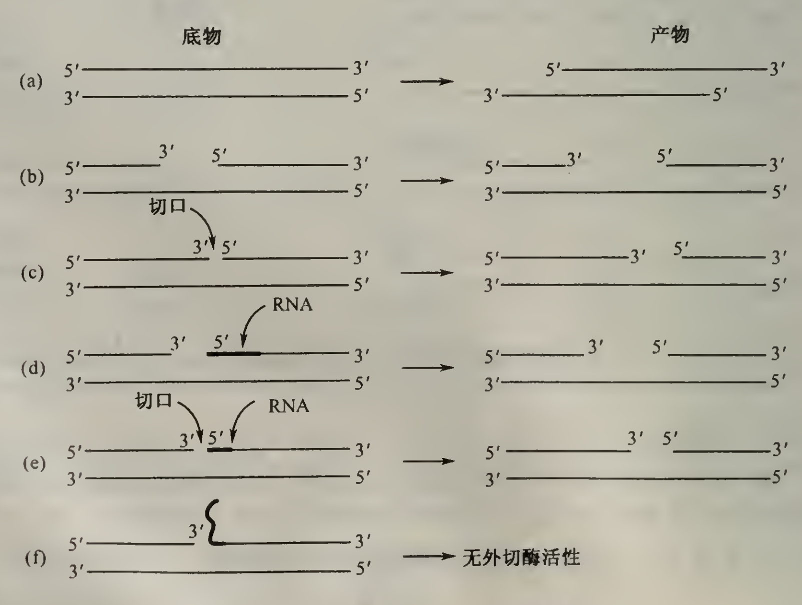


图 4-6　DNA 聚合酶 I 5′→3′外切核酸酶活性

末端核苷酸，不切除不配对的单链 5′-P 末端核苷酸；④既能切除脱氧核苷酸也能切除核苷酸；⑤对只具有 5′-P 末端的切口也有活性。

DNA 聚合酶 I 的 5′→3′外切核酸酶活性和聚合活性同时作用可以进行切口翻译（nick translation），可用来制造放射性探针。在反应物里加入同位素标记的脱氧单核苷酸后，DNA 聚合酶 I 首先利用 5′→3′外切核酸酶活性逐个切下 DNA 上的脱氧核苷酸，再逐个将标记的脱氧核苷酸加上去（图 4-7）。

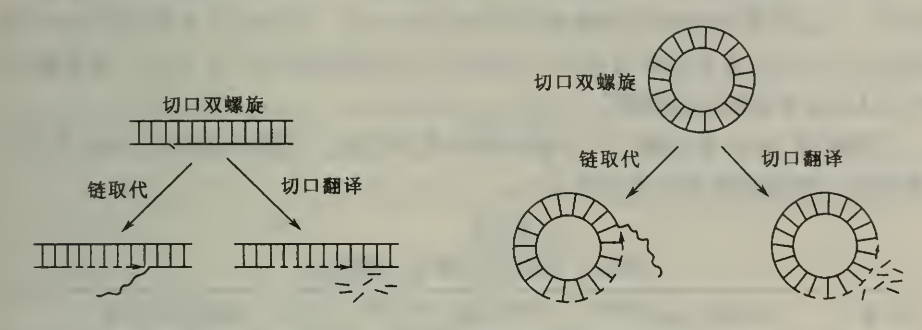


图 4-7　链取代和切口翻译

控制条件使 DNA 聚合酶 I 在缺口处只具有聚合活性不具外切核酸酶活性，使新生的链取代亲代链（strand displacement）。链取代是 DNA 重组中的一个重要步骤。在体外有一种现象常常伴随着链取代反应发生，叫模板开关（template-switching），但还未发现该反应的生物学意义。

（4）内切核酸酶活性

DNA 聚合酶 I 具有 5′→3′外切核酸酶活性，也具有 5′→3′内切核酸酶活性，如图 4-8 所示。在两个碱基之间切开从而产生一个具有 5′-P 末端的不配对碱基的片段。

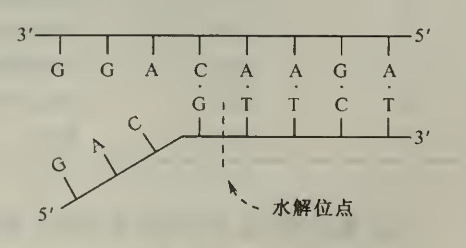


图 4-8　DNA 聚合酶 I 的内切核酸酶活性

很多实验证明 DNA 聚合酶 I 不是大肠杆菌复制中主要的合成酶，而是在除去 RNA 引物和损伤修复中起重要作用。

2. 大肠杆菌的 DNA 聚合酶 II

纯化的大肠杆菌 DNA 聚合酶 II 分子质量为 90 000～120 000Da，由一条肽链构成。它从 5′→3′方向合成 DNA，当四种脱氧核苷酸、Mg^{2+}、NH_4^+ 和一个带 50～200 碱基缺口的双链 DNA 存在时达到最大活力，当缺口加大时反应速度降低，但加入 DNA 结合蛋白后能恢复到原来的活力，该酶也需要带 3′-OH 末端的引物，具有 3′→5′外切核酸酶活性，但没有 5′→3′外切核酸酶活性。DNA 聚

合酶 II 的功能是在应急修复（SOS）应答中起作用，这个酶也可与 DNA 聚合酶 I 一样填补缺口，并且在 DNA 的损伤修复中更容易利用损伤的模板直接合成 DNA。

3. 大肠杆菌 DNA 聚合酶 III

带有 DNA 聚合酶 III 的温度敏感突变（polC）的大肠杆菌在限制温度下不能存活，从这种菌株得到的裂解液也不能合成 DNA，当加入正常细菌的 DNA 聚合酶 III 时可以恢复它的聚合能力，因此与 DNA 聚合酶 I 和 II 不同，聚合酶 III 是体内 DNA 复制所必需的酶。

大肠杆菌 DNA 聚合酶 III 全酶由多种亚基组成，组成全酶的各种亚基及其功能以及不同装配体的符号见表 4-2。

表 4-2　DNA 聚合酶 III 全酶组成

亚 基	分子量（kDa）	功能	亚装配体及符号			
α	129.9	DNA 聚合酶				
ε	27.5	3′→5′外切核酸酶	核心	pol III*		
θ	8.6	促进 ε 外切核酸酶				
τ	71.1	核心酶二聚化				
		结合 γ 复合物				
γ	47.5	结合 ATP			pol III*	pol III 全酶
δ	38.7	与 β 结合	γ 复合物			
δ′	36.9	与 γ 和 δ 结合	（依赖 DNA 的			
χ	16.6	与 SSB 结合	ATPase）			
ψ	15.2	与 χ 和 γ 结合				
β	40.6	滑动				

DNA 聚合酶 III 与 DNA 聚合酶 I 类似，有聚合活性以及 3′→5′和 5′→3′的外切核酸酶活性，DNA 聚合酶 III 的作用范围如图 4-9 所示。pol III 的 5′→3′外切核酸酶活性与 pol I 的不同，pol III 的 5′→3′外切核酸酶活性只对单链 DNA 有作用，因此不能用于切口翻译。

DNA 聚合酶 III 全酶在复制中十分重要，它可以在 DNA 模板引物的 3′-OH 上以每分钟约 60 000 核苷酸的速率逐个延伸新生的 DNA，三种酶的性质比较见表 4-3。

4. 真核生物中的 DNA 聚合酶

所有真核生物中的聚合酶的性质基本上同大肠杆菌中的聚合酶相似，它的活性依赖于模板，带 3′-OH 的引物，四种脱氧核糖核苷三磷酸。动物细胞的 DNA

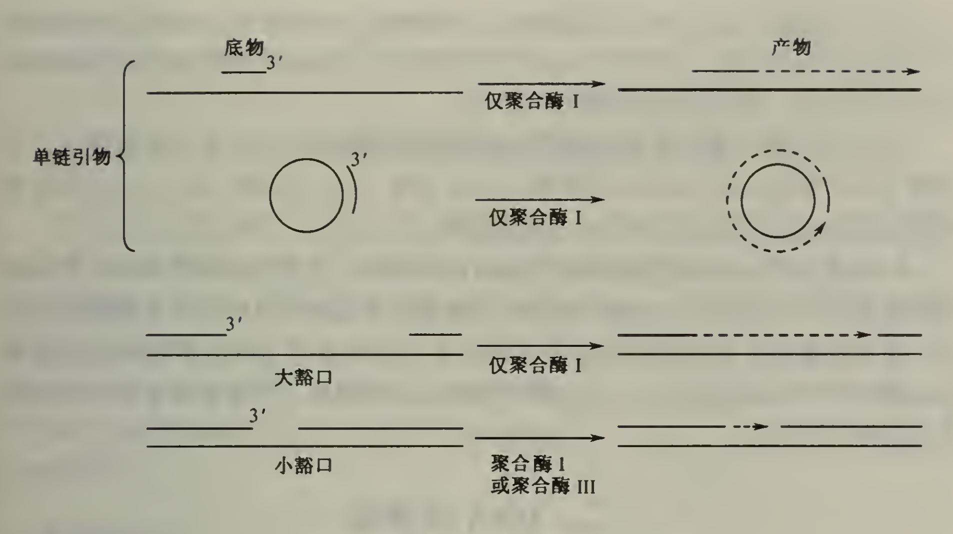


图 4-9　聚合酶 I 和聚合酶 III 的活性

DNA 聚合酶 III 只能在一个小的豁口处起作用。

表 4-3　三种大肠杆菌 DNA 聚合酶性质的比较

	pol I	pol II	pol III
功能：5′→3′聚合	+	+	+
3′→5′外切	+	+	+
5′→3′外切	+	−	+
焦磷酸化和 PPi 交换	+		+
模板－引物：			
完整双链	−	−	
有引物的双链	+	−	−
有断口的双链（多聚 dAT）	+	−	−
有缺口的双链或 5′端突出的单链：			−
单链长度＜100 核苷酸	+	+	+
单链长度＞100 核苷酸	+	−	−
分子质量/1000U	109	120	140
每个菌体中的分子数（估计值）	400		10～20
结构基因	*polA*	*polB*	*polC*

聚合酶活性分布在几个不同的细胞级分中，至今只对可溶性的 DNA 聚合酶研究得比较多。DNA 聚合酶 α 在迅速增殖的细胞中活力较高，这个酶的分子质量在 165 000～175 000Da 之间。它的最合适的底物是带缺口的双链 DNA，带引物的变性 DNA 也是很好的底物。

DNA 聚合酶 β 是一个分子质量小（43 000Da）的聚合酶，在整个细胞周期中它的活力没有变化，它以 DNaseI 处理过的天然 DNA 作为模板，对变性 DNA 几乎没有活性，也没有外切核酸酶的活性。

DNA 聚合酶 γ 是一个大分子酶，在细胞内含量很低，因此不容易纯化。这个酶可以以带缺口的 DNA 作为底物，但以多聚（A）、寡聚（dT）$_{8\sim10}$作为底物时酶活力更高，因此它更像是一个反转录酶。

DNA 聚合酶 α 可能是真核生物 DNA 的复制酶，主要根据是聚合酶 α 的活性随细胞周期变化而变化。如淋巴细胞中聚合酶 α 的活力和 DNA 合成周期是一致的，而聚合酶 β 活力的升高要后随一段时间。在 Hela 细胞中，聚合酶 α 的活性从 G 期开始就升高 10 倍，一旦 S 期结束酶活力就降低，而聚合酶 β 的活性在整个过程中没有变化。

二、DNA 连接酶

DNA 连接酶（DNA ligase）催化一条 DNA 链的 5′磷酸基团与另一条 DNA 链的 3′羟基形成磷酸二酯键，但是这两条链都必须同时与另外一条链互补结合，而且两链必须相邻，即该酶只能连接一个切口（nick）而不能连接一个缺口(gap)，如图 4-10 所示。连接酶所需要的能量来自 NAD（如大肠杆菌）或 ATP (如 T4 诱导的连接酶和动物细胞中的连接酶)。T4 诱导的连接酶不仅能在模板上连接 DNA 和 DNA 之间的切口，而且能连接 RNA 和 RNA 之间的切口；不仅能连接 DNA 中的单链切口或双链的黏性末端，而且也能连接无单链黏性末端的平头双链 DNA。大肠杆菌连接酶则不行。

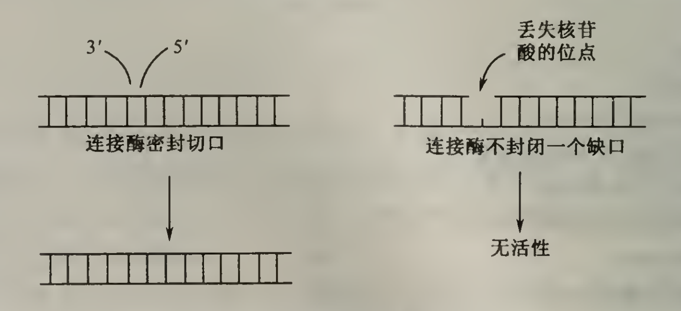


图 4-10　DNA 连接酶的作用

只连接一个切口，不能连接一个缺口。

DNA 连接酶在 DNA 复制、重组和修复中的作用是明显的，缺乏 DNA 连接酶的大肠杆菌突变株中会发生冈崎片段积累，而且对紫外光敏感性增加。每个大肠杆菌约有 300 个连接酶分子，在 30℃ 每分钟能连接 7500 个切口，而在实际的

生命活动过程中每分钟大约只有 200 个切口需要连接，因此连接酶的含量对于 DNA 的复制而言是足够的。

三、与解链有关的酶和蛋白质

DNA 是双螺旋链，在复制过程中双螺旋要解开，双螺旋的解开使得复制叉前面产生正超螺旋，这些应力如果不以某种方式释放将妨碍复制叉的前进。如果让这些超螺旋传到分子末端才释放的话，对大肠杆菌而言，DNA 分子必须每分钟旋转 20 000 转左右，而且环状分子还不能以这种方式释放正超螺旋，除非分子内有一个切口使链可以自由转动。现在已经找到一些酶和蛋白质，它们或者能使 DNA 双链变得容易解开，或者可以使超螺旋分子松弛。以下介绍一些这种类型的蛋白质。

1. 单链结合蛋白

最初发现噬菌体 T4 的基因 *32* 编码的蛋白质可以在远低于 T_m 的温度下使双链 DNA 拆开并牢固地结合在单链 DNA 上，后来发现许多种生物中都有这类蛋白质并将其定名为单链结合蛋白（single strand binding protein，SSB）。在文献中这类蛋白质曾用过多种名称，如解旋蛋白（unwinding protein）、熔化蛋白（melting protein）、螺旋降稳蛋白（helix destabizing protein）。

从原核生物中得到的 SSB 在与 DNA 结合时表现出协同效应。比如第一个 SSB 结合到 DNA 上去的能力为 1，第二个 SSB 结合的能力就会高达 10^3。这可能是由于蛋白质和蛋白质之间的互相作用增强了结合能力；也可能由于第一个 SSB 的结合使 DNA 链的结构发生了变化，有利于后来的 SSB 与 DNA 的结合；很有可能两个假设原因都存在。

但从真核生物如牛、鼠和人细胞中得到的 SSB 蛋白与单链 DNA 结合时不表现为协同作用。

SSB 蛋白能刺激同源的 DNA 聚合酶的活力，如 T4 的基因 *32* 编码的蛋白能刺激 T4 DNA 聚合酶活力，但不能刺激大肠杆菌 DNA 聚合酶的活力，反之亦然。

SSB 蛋白能使天然 DNA 的熔点降低，这可能不是它在生物体内的唯一功能。单链 DNA 与 SSB 蛋白结合之后能抵抗核酸水解酶的作用，如大肠杆菌 SSB 蛋白能保护单链 DNA 不受脱氧核糖核酸酶、蛇毒磷酸二酯酶、S1 核酸酶、红色链孢霉内切核酸酶以及 DNA 聚合酶 I 和 III 的 3′→5′外切核酸酶的降解，但对小球菌核酸酶和大肠杆菌外切核酸酶 I 仍然敏感，因此 SSB 蛋白的另一个重要生物功能可能就在于保护复制中的 DNA 单链部分不被核酸酶降解。

真核生物的 SSB 蛋白和原核生物的 SSB 蛋白可能有些不同，原核生物的

SSB 蛋白同双链 DNA 的结合能力很弱，而真核生物的 SSB 蛋白与双链 DNA 有一定结合能力；原核生物的 SSB 蛋白能促使变性的 DNA 复性，而真核生物的 SSB 蛋白没有这种能力。因此原核生物 SSB 蛋白的作用方式是先结合已存在的单链区从而侵蚀双链部分，而真核生物 SSB 蛋白是先结合 DNA 双链然后使之熔解。

2. 解链酶

通过水解 ATP 获得能量来解开双链 DNA 的酶称为 DNA 解链酶。

解链酶水解 ATP 的活力依赖于单链 DNA 的存在，如果双链 DNA 中有单链末端或缺口，则 DNA解链酶可以首先结合在这一部分，然后逐步向双链方向移动。复制时大部分 DNA 解链酶可以沿着后随链模板的 5′→3′方向随着复制叉的前进而移动，只有 rep 蛋白是沿着前导链模板的 3′→5′方向移动，因此在复制时很可能是 rep 蛋白和某种 DNA 解链酶分别在 DNA 的两条母链上协同作用以解开双链 DNA（图 4-11）。

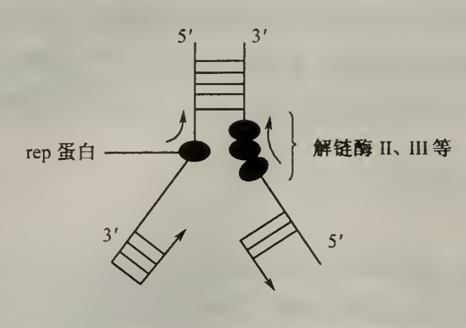


图 4-11　解链蛋白和 rep 蛋白的协同作用

已经发现，有些 DNA 解链酶同时还有引发酶的功能。例如 T7 的基因 *4* 编码的蛋白质可以优先与单链 DNA 结合并在母链的 5′→3′方向靠水解 NTP 而移动并解开双链，当它到达 DNA 母链的某一识别位置时就可以按模板的要求合成小片段 RNA 作为后随链的引物，比较多的引物链是 pppACCC 或 pppACCA。

3. 拓扑异构酶 I

最初从大肠杆菌中分离到的一种叫 ω 蛋白的蛋白质可以将共价闭环的超螺旋DNA 变成松弛态而不改变其化学组成或结构。由于反应的产物还是闭环 DNA，因此酶反应机制只可能是先切开双链 DNA 中的一条链，使链的末端沿螺旋的方向转动，然后再把切口封起来。后来发现这种酶广泛存在于各种生物中并将其命名为拓扑异构酶 I（topo I），不同学者曾给予不同的命名，如转轴酶(swivelase)、松旋酶（untwisting enzyme)、DNA 松弛酶（DNA relaxing enzyme)、切割封口酶（nicking-closing enzyme)。

Topo I 在松弛 DNA 超螺旋时不需要 ATP，它可以催化三种拓扑转换反应(图 4-12)。

Topo I 在复制中的作用还不十分清楚，某一大肠杆菌的突变株缺失 Topo I 酶基因，但它的生长速度只比正常细菌稍慢一点，因此它并不是细菌生长必不可少的酶。

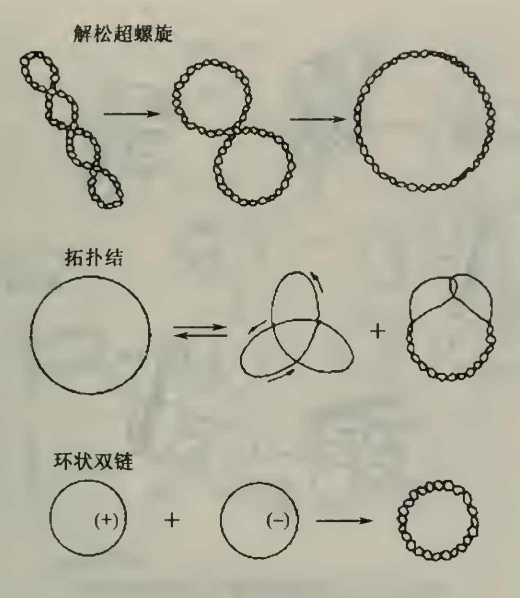


图 4-12　拓扑异构酶 I 的作用

4. 拓扑异构酶 II

拓扑异构酶 II 又称为 DNA 解旋酶（DNA gyrase）或紧旋拓扑异构酶（twisting topoisomerase），它由两个 α 亚基和两个 β 亚基组成。

拓扑异构酶 II 可以在水解 ATP 的同时使松弛状态的环状 DNA 转变为负超螺旋 DNA，其反应机制不同于拓扑异构酶 I，每次引起的超螺旋变化数为 2（图 4-13）。这种反应说明拓扑异构酶 II 每次能使两条 DNA 链同时切断，让其中一条双链 DNA 穿过断口，再将已切断的末端重新封接起来，其反应过程可参考图 4-14。

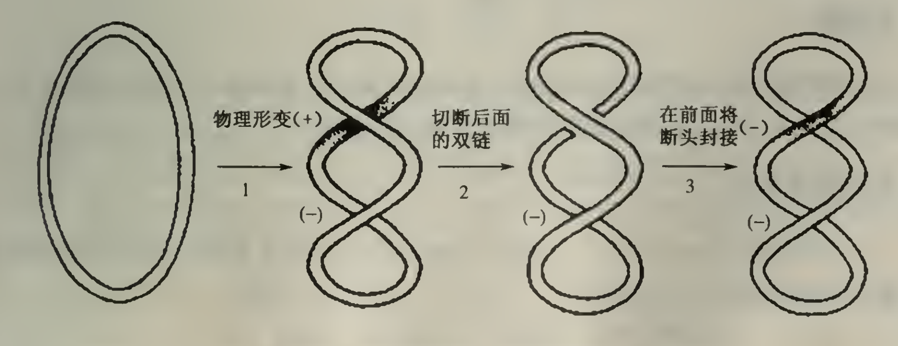


图 4-13　拓扑异构酶 II 符号变换反应

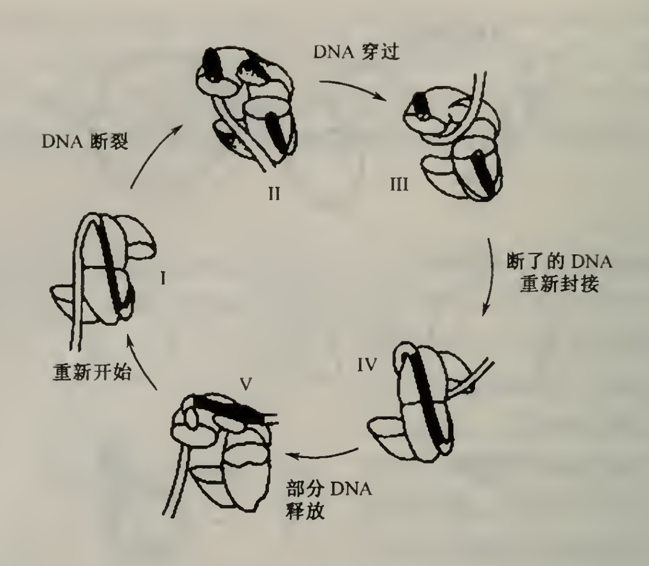


图 4-14　拓扑异构酶 II 的作用机制

拓扑异构酶不能非常专一地识别 DNA 序列，但常见的识别结构是

5′—TAT↓　GNTN　NT—
—ATA　CNAN↑　NA—5′

在加工过程中两条链也不是平行的（临时切断处用箭头表示），而且两条链之间相差四个核苷酸。

Topo II 在没有 ATP 时可以使超螺旋 DNA 变为松弛态。

Topo II 广泛存在于各种生物中，它是完成复制所必需的一种酶，至少可以分开完成了复制的两个子代环状双链 DNA。图 4-15 表示 SV40 DNA 在 Topo II 作用下的复制及其子代 DNA 的分开。

5. 引物酶

大肠杆菌中的引物酶（primase）是 DnaG 蛋白，它催化在 DNA 复制中 RNA 引物合成，是 DNA 复制起始复合物中的复合蛋白之一。

6. 引发前体蛋白

引发前体蛋白 n、n′、n″、DnaB、DanC 和 i 是 DNA 复制起始过程中装配成引发前体的蛋白质。

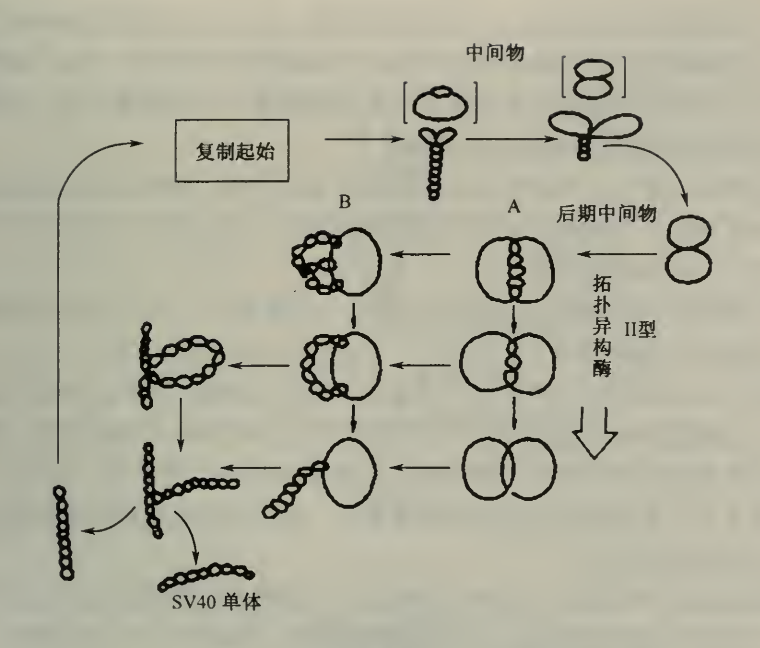


图 4-15　拓扑异构酶 II 的功能示意图
垂直箭头表示在 Topo II 作用下逐渐拓扑异构化的过程，
横向箭头表示 DNA 合成，最后得到两个分开的超螺旋单体。

第三节　DNA 复制的起始、延伸和终止

一个 DNA 分子的复制当然不是瞬间就能完成的，它是一个在空间上井然有序的复杂的生化反应过程。DNA 复制是以复制子为单位进行的。

一、复　制　子

DNA 分子上一个独立的复制单位叫做复制子（replicon）。一个复制子包括一个复制原点（origin），也可能包括一个复制终止点（terminus）。

任何一个与复制原点连接而又不被复制终止点隔开的 DNA 序列都作为该复制子的一部分而被复制。复制原点是一个顺式作用位点，它只能影响它所在的那条 DNA 分子。

在原核生物中，复制子被看成是一个具有一个复制原点和一个编码复制调控蛋白基因的单位，而用于真核生物染色体上的复制子的一般描述为：含有一个复制原点的复制单位，其反式作用调控蛋白可能由其他位置编码（即调控蛋白基因不包括在复制子中）。

一个原核细胞的基因组由一个复制子组成，所以 DNA 复制和细胞分裂同时发生。最大的复制子是大肠杆菌基因组，一旦细胞分裂开始，DNA 复制也就从一个单一的复制原点起始并完成整个染色体的复制。因为细菌只有一条染色体，所以这种复制控制类型叫做单拷贝控制。

有些细菌中除自己的染色体以外还可能含有质粒，质粒有自己的复制原点，是自主复制单位。质粒复制子可以与细菌染色体一起复制，此时是单拷贝控制；但质粒也可以是多拷贝控制。

每个噬菌体或病毒的 DNA 也可以组成一个复制子，在一个感染周期中，其 DNA 的复制可以起始多次。

复制子是个灵活可变的单位。细菌基因组中，DNA 分子以半保留复制产生复制品，而噬菌体或质粒则可以产生单链单体或多聚体形式的基因组。复制子的复制模式依赖于在复制原点上起始复制的那些相互作用的性质，复制一旦开始，就会继续下去，直至整个基因组全部被复制。复制起始的频率由调控蛋白和复制原点的互相作用来决定。

真核基因组中含有多个复制子，所以一个分离单位就包括许多复制子，这就增加了其控制问题的复杂性。在一个细胞周期中，染色体上的所有复制子必须几乎同时被激活，而且在同一个周期中只被激活一次。

必须存在一些信号去区别复制过的复制子或没有复制过的复制子，以便复制子在同一周期中决不会第二次被起动复制。因为许多复制子是单独被激活的，所以还必须存在一些信号来指示所有复制子的复制全过程什么时候已经完成。

目前收集到的有关复制子构建和有关复制子之间关系的资料还极少，还不知道在每个细胞周期中复制的途径是否相同，也不知道是否所有的复制原点总是被使用，还是有些复制原点有时是沉默的，更不知道复制原点是否总是以同样的次序起始。如果存在不同类型的复制原点，那么又由什么特征来区别它们呢?

叶绿体和线粒体 DNA 复制的调控更像在一个细菌细胞中存在的多个拷贝的质粒那样，细胞中的每个细胞器 DNA 都有多个拷贝，而且细胞器 DNA 的复制调控必须和细胞周期相关联。在所有的系统中，关键的问题是解释作为复制原点的功能序列如何被复制装置中合适的蛋白质所识别。

用同位素标记方法可以研究复制子的复制起始点以及复制方向，也可以研究双向复制中复制速度是否相同。应用此技术已证明 DNA 复制原点最常用于双向复制（图 4-16)。

细菌基因组是单个复制子，一个细菌基因组要遗传，其复制必须满足几个条件：①一个复制周期的起始；②对起始频率的控制；③复制后的染色体分配到子细胞。

前两个条件由复制原点来实现，分离则可能是一个独立的功能，但事实上，常常由复制原点邻近的序列来实现（真核生物的复制原点在分离中绝对没有作

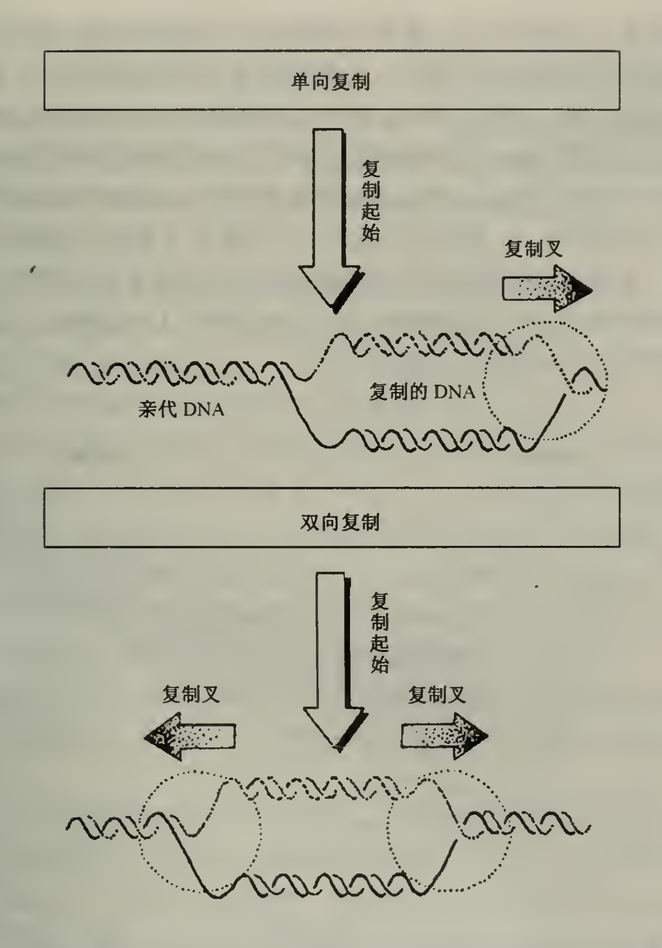


图 4-16　复制子可能是单向或双向，这取决于在起始区是形成一个还是两个复制叉

用)。用缺失突变的方法我们可以鉴定出每一个与分离有关的序列。

一个复制原点连接到任一 DNA 分子上都可以支持其复制，因而可以以此来鉴定复制原点。

在真核细胞中，DNA 复制属于细胞周期的一部分。在高等真核生物中，细胞周期的 S 期占据间期且常常持续几个小时，通过许多独立的复制子的复制而完成真核生物染色体 DNA 的复制。

细菌中，DNA 复制与细胞生长相协调。首先是调整复制周期的起始频率使之与细胞生长的速率相适应，其次是复制周期的完成与细胞分裂相联系。

二、半保留复制和半不连续复制

在细胞分裂过程中，亲代细胞所含有的遗传信息原原本本地传递给两个子代

细胞。我们知道，细胞的 DNA 携带了细胞的一切遗传信息，因此，亲代 DNA 也必须准确无误地复制成两个拷贝，并分配到两个子代细胞中去。研究 DNA 复制的目的就是要了解：①子代 DNA 为什么能够真实地获得亲代 DNA 的遗传信息；②复制是怎样进行的；③生物体是怎样对 DNA 复制进行调控的。

关于前两个问题，Watson 和 Crick 在发表 DNA 双螺旋结构模型时就讨论过了。在 DNA 双螺旋中，G 只能和 C 配对，A 只能和 T 配对，因此依照碱基配对原则，任意一条链的碱基序列都可以精确地决定它的互补链的序列。在 DNA 复制时，只要简单地把 DNA 双链拆开，每一条链都可以作为模板，从而合成结构上完全一样的两条 DNA 双链（图 4-17）。

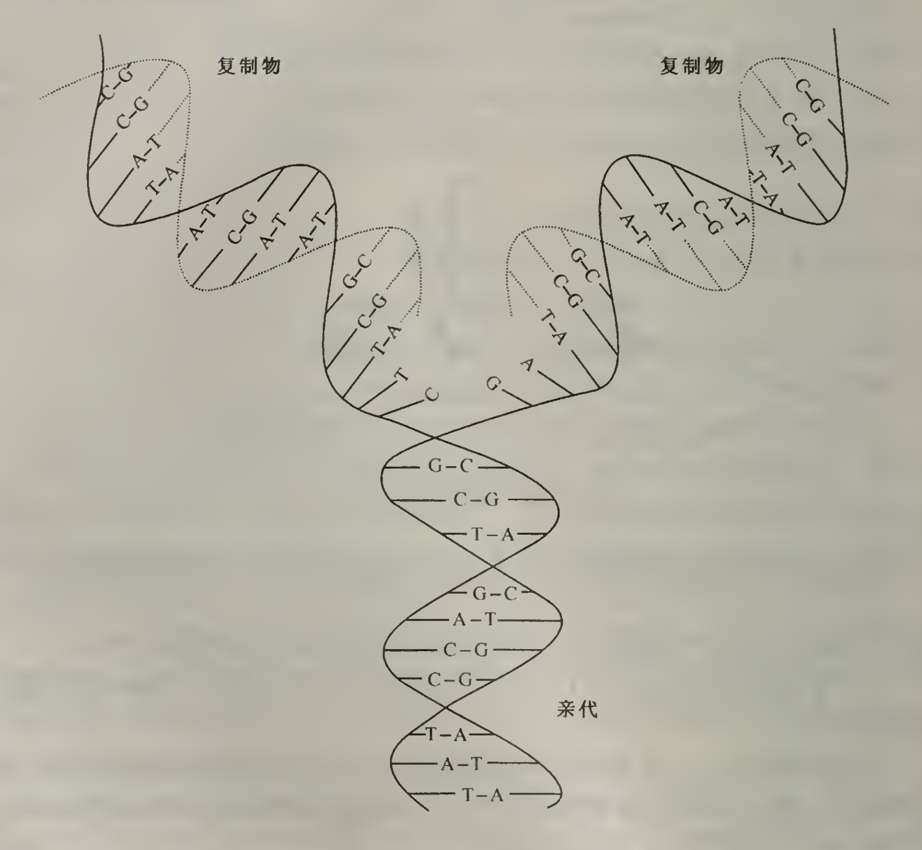


图 4-17　Watson 和 Crick 提出的 DNA 复制机制

然而，这样一个表面上看来很简单的过程其实并不像最初提出时那样，可以像拉“拉链”一样很容易完成而不需要任何酶的帮助。实际上，复制过程十分复杂，它涉及许多酶和蛋白质的协同作用。目前虽然对大部分与复制有关的酶和蛋白质的结构和性质已研究清楚，但对于 DNA 的双链究竟是如何解开以及怎样保

持其复制的真实性的问题仍是知其然而不知其所以然。以下事实足以表明复制过程的复杂性：① 需要为解螺旋提供能量；② 单链 DNA 倾向于链内碱基配对；③ 一种酶只能催化有限的生理生化反应；④ 一系列的防护措施既能防止复制错误的产生又能消除偶然出现的错误；⑤ DNA 分子均以高度压缩状态和环状超螺旋状态存在，必须解决由此给复制系统带来的强大的几何强制力。这些问题一直是人们力求弄清楚的问题。

Watson 和 Crick 提出的 DNA 复制方式称为半保留的复制方式。在理论上还可以设想其他一些复制方式，例如全保留复制和散布式复制。

当半保留复制的假设提出时，人们并不了解 DNA 能变性而且认为链分离是不可能的；链不可分离的思想主要由于对解螺旋所需要的时间的不正确计算（值很大）和对 DNA 变性后得到分子质量不等的 DNA 实验的错误解释。此外，一个细菌细胞的 DNA 的长度大约是 1.3mm，而一个典型的杆菌的长度只有 0.003mm，所以认为两条链完全解开进行链分离的想法几乎是荒谬的。

全保留复制的模型认为，在复制前后 DNA 都是缠绕在一起的，仅仅是复制位点解开几个碱基便足以使多聚酶阅读并产生新的分子，而亲代分子保持完整。因此在全保留复制中，亲代 DNA 分子完全不需要解链。

散布式复制模型认为亲代 DNA 分子先打成片段然后再互为模板复制。于是亲代 DNA 片段分散在子代 DNA 分子中。这个模型可以避开半保留复制中的一些问题。

然而以下几个著名实验证明了 DNA 分子以半保留方式进行复制。

Meselson-Stahl 实验：1958 年 Meselson 和 Stahl 设计了一个很巧妙的实验，证明 DNA 复制是以半保留方式进行的。他们先把大肠杆菌在 96.5%同位素纯度的$^{15}NH_4Cl$ 培养液中培养 15 代，经过多次细胞分裂使所有细菌 DNA 都被^{15}N 标记。然后把这些细胞转移到正常的$^{14}NH_4Cl$ 培养液中培养。在不同时间取出样品，用 SDS 裂解细胞。裂解液进行 CsCl 密度梯度离心，达到平衡时，DNA 分子就停在等密度区不再移动，紫外光下可以观察这条吸收带。含有^{15}N 的 DNA 分子密度在 1.71g/cm^3 左右，大于^{14}N 的 DNA 密度，它在离心管较低的部位形成一条带。让这种重标记的 DNA 在^{14}N 培养液中复制一代，发现重密度带消失了，而在比正常 DNA 稍低的地方形成另一个带，我们称它为中间密度带。在^{14}N 培养液中再复制一代，就可看见在 CsCl 密度梯度溶液中出现两条带：一条在中间密度区域，另一条在轻密度区域，而重密度区域仍然无带。随着在^{14}N 培养液中培养的代数的增加，中间密度区带的量逐渐减少，低密度区带的量逐渐增加，但并不出现新的密度区带。这个实验结果（图 4-18），证明 DNA 分子半保留复制理论是正确的，排除了全保留复制方式的可能性；因为全保留复制方式中亲代 DNA 毫无变化地转移到子代中，那么重密度的 DNA 带不会消失。在半保留复制中，DNA 是密度杂交分子，两条链中一条链是亲代^{15}N 标记的重链，另一条是

含^{14}N的轻链，这样两条链形成密度上杂交的DNA。因此所有子代DNA的密度都比亲代DNA的密度小，但又比正常DNA的密度大，形成一个中间密度带。如果把这种密度杂交的分子在^{14}N培养液中再培养一代，按半保留复制方式就会形成$^{15}N^{14}N$和$^{14}N^{14}N$两种分子，所以形成两条带；而且随着在^{14}N培养液中复制代数的增加，$^{15}N^{14}N$ DNA相对量减少，而$^{14}N^{14}N$ DNA相对量增加，但不出现新带，就如图4-18所示。

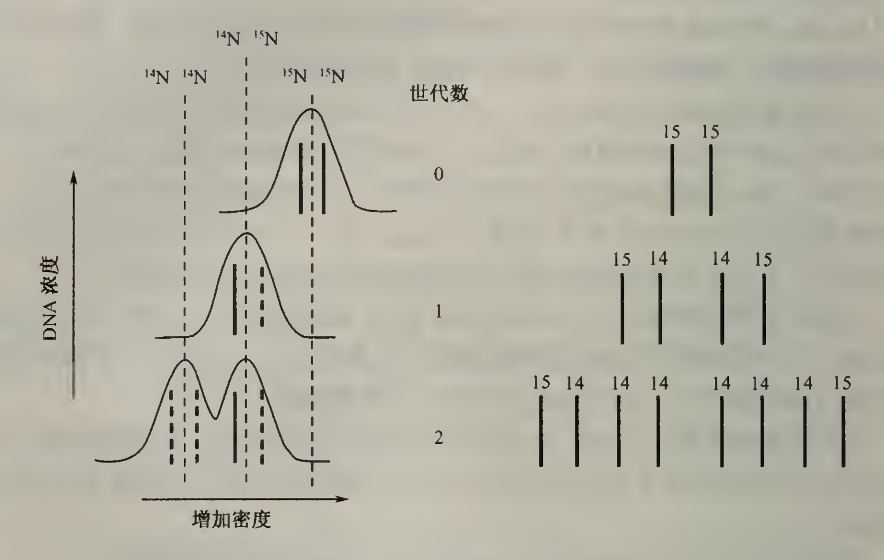


图4-18　*E.coli* ^{15}N培养基中生长几代后转入^{14}N培养基中，培养一代之后，全部DNA呈杂交密度。右边是不同时间分离的DNA的状态

第二个实验是：把$^{15}N^{14}N$密度杂交DNA加热至100℃变性，使两条链分开，然后在CsCl密度梯度中离心，变性的密度杂交DNA离心后出现了两条带，即^{14}N-DNA和^{15}N-DNA带，结果如图4-19所示。这说明在^{14}N培养液中第一代复制的DNA是由一条^{14}N轻链和一条^{15}N重链组成的，这就进一步确证了DNA的复制是半保留方式。

1963年Cairns用放射自显影的方法阐明大肠杆菌DNA是环状分子并且以半保留方式进行复制。1957年Taylor用^{3}H脱氧胸苷的溶液标记生长中的蚕豆苗证明了真核生物的染色体复制也是以半保留方式进行的。

以上实验得出一个共同的结论，DNA是以半保留方式进行复制的，无论是原核还是真核都是如此。

DNA半保留复制已经确立，现在我们面临的问题是如何解释解螺旋。下面将讨论这个问题。

在半保留复制中双螺旋的解旋存在一个机制问题。图4-17中在复制叉处两条链必须互相绕着旋转。如果DNA分子在细胞里是伸展的，事情还简单些；然

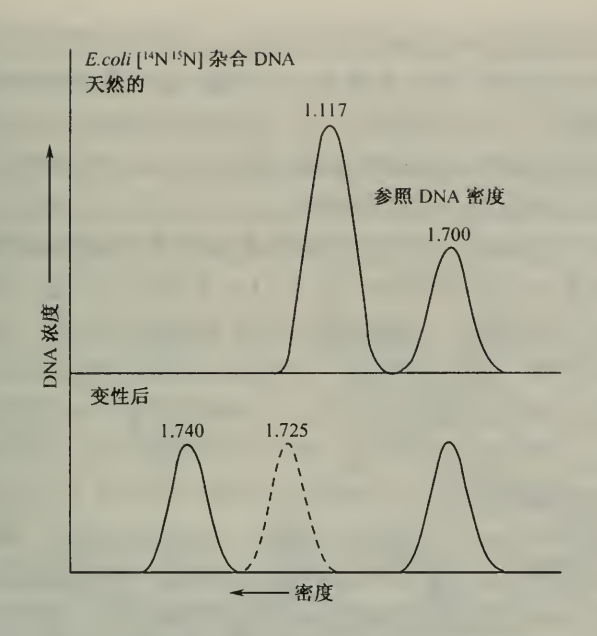


图 4-19 $^{15}N^{14}N$ 双股 DNA 分子和单股多核苷酸链的密度

1.700 是参照 DNA 的密度，下部是$^{15}N^{14}N$DNA 变性后，和参照 DNA 一起密度梯度离心的结果。

而 DNA 分子的实际长度是它在细胞中长度的 600 倍，这样一个高度被压缩的分子要进行旋转是不可能的。解决这个问题的简单方法是在复制叉前面的亲代链上造成一个切口，这样使得超螺旋区的一个小片段发生旋转，从而就解决了它的几何学问题（图 4-20）。在此之后还需要被打开的链重新连接，而且这个切口的修复必须在复制叉到达这个切口之前完成，否则子代链就会丢失（图 4-20）。细胞内存在相应的酶和蛋白质能切开 DNA 单链并且能把切口再封起来。

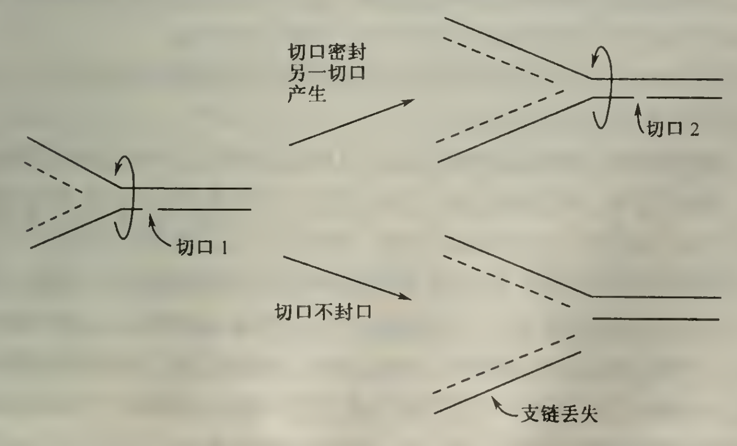


图 4-20 复制叉前边产生切口以利旋转，如果切口不被封上，新形成的子代链将丢失

因为 DNA 双螺旋的两条链是反向平行的，因此在复制叉附近解开的 DNA 二条链一条是 5′→3′方向，另一条是 3′→5′方向。由于模板的极性不同，是不是一种 DNA 聚合酶向 5′→3′方向延伸，另一种 DNA 聚合酶可以沿 3′→5′方向延伸呢？至今发现的所有体内 DNA 聚合酶都只能以 5′→3′方向延伸。而且由于复制真实性的要求，也只能以 5′→3′方向延伸。

1968 年冈崎（Okazaki）发现：大肠杆菌用^{3}H 脱氧胸苷短时间标记的 DNA 变性后，用超速离心方法可以得到许多 8～10S 的标记小片段，这些小片段被称为冈崎片段。标记时间延长，冈崎片段可转变为成熟 DNA 链，因此它们必然是复制过程的中间产物。按照这个结果，冈崎认为 DNA 合成至少在一条链上是不连续的。以复制叉向前移动的方向为准，一条模板链是 3′→5′方向，称为前导链（leading strand）。原则上在前导链上，DNA 的合成能以 5′→3′方向（指新合成的 DNA 链的极性）连续合成。另一条链上新生 DNA 也是以 5′→3′方向合成，但与复制叉前进的方向相反，这一条链称后随链（lagging strand）。随着复制叉的移动，后随链上不连续的 DNA 片段可以连接起来成为一条完整的 DNA 链。

此外发现在 DNA 连接酶的突变株中积累了冈崎片段；利用电镜技术直接观察到复制叉的一边是双链，另一边是单链，或一段很短的双链。这些实验也证明复制过程中出现冈崎片段。这种前导链的连续复制和后随链的不连续复制在生物界是普遍存在的，称为双螺旋的半不连续复制。很多因素，如模板的损伤，复制蛋白或 dNTP 供应不平衡等都可以影响前导链的连续复制，引起前导链从另一新起点开始。

三、不同的复制引发机制

DNA 复制过程中需要合成 RNA 引物，这对合成后随链也同样适用。RNA 引物的合成称为引发。后随链的引发是个十分复杂的过程。

现在知道大肠杆菌 DNA 的复制需要多种蛋白质和酶的协同作用，其中的不少蛋白质都与引发有关（表 4-1）。

由于大肠杆菌的引发机制过于复杂，人们首先研究了大肠杆菌噬菌体的引发过程。在大肠杆菌中，噬菌体 M13 DNA 可以利用细菌的 RNA 聚合酶，先在 M13 单链 DNA 上合成一小段 RNA 引物，然后在 DNA 聚合酶的作用下，以 M13 为模板延伸新生的 DNA 链。但当存在利福平时，这种单链 M13 DNA 变为双链 DNA 的反应就完全被抑制了，因为利福平可以抑制大肠杆菌 RNA 聚合酶的活力。但是在一种大肠杆菌突变株中，M13 DNA 的合成不受利福平的抑制；而且单链 *ΦX*174 DNA 变为双链的 RF 型也完全不受利福平的影响，因此可能存在一种以前未认识到的 DNA 合成系统。深入研究的结果导致了引发体（primosome）的发现。现在已经证明无论是原核还是真核生物，DNA 复制时都要靠引发体来

引发。引发体包括引发前体（preprimosome）和引发酶（pimase)。引发体像火车头一样在后随链上向分叉的方向前进，并在模板上断断续续地引发生成后随链的引物 RNA 短链。

复制叉处有两个 *E.coli* DNA 聚合酶 III 核心酶，在前导链上 pol III 沿着 5′→3′滑动，在后随链上后随链先弯曲。使得 pol III 移动方向与前导链上的合成方向一致，并使前导链的合成的速度不比后随链快得太多（图 4-21)。

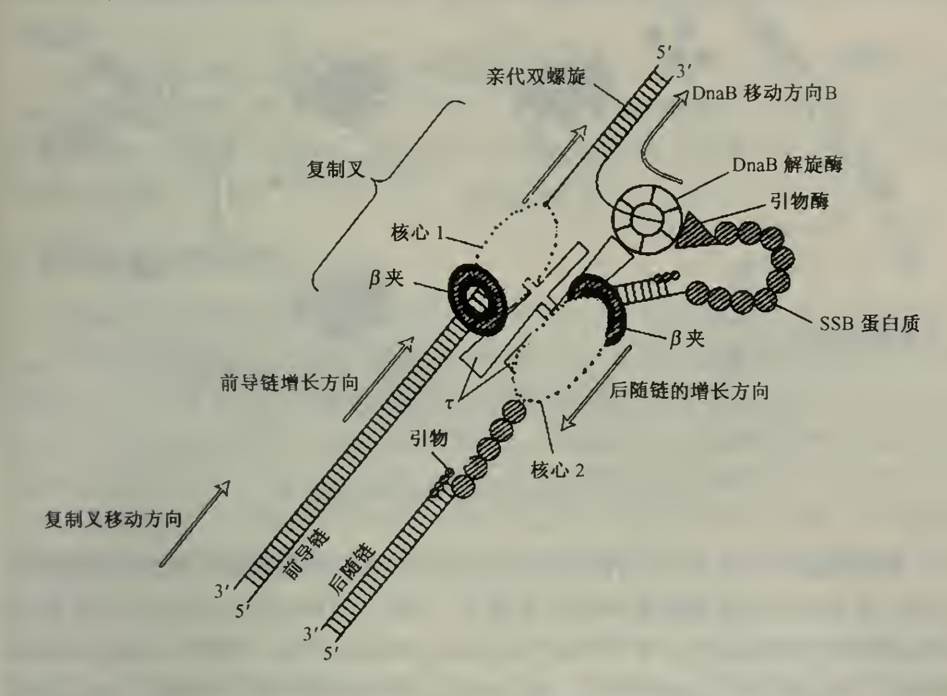


图 4-21　在增长的复制叉处 *E.coli* 复制蛋白之间的关系示意模型

尽管 DNA 的复制要靠引发体来引发，然而不同形式的 DNA 复制的引发及复制形式却不尽相同。生物界中，许多生命形式的遗传物质是环状双螺旋分子。下边将以几种生物为例来了解它们的 DNA 复制的引发机制及复制形式。

1．大肠杆菌 DNA 复制的起始

前边已讨论了复制原点的结构特点和用于 DNA 复制的酶及蛋白质，现在来讨论这些酶和蛋白质是如何相互作用而起始 DNA 复制的。

首先大肠杆菌的 DnaA 蛋白结合到 oriC 的 9 聚体（9-mers，9bp）上，有利于 DNA 双螺旋的熔解和链分离，熔链的位置发生在 oriC 的 13 聚体（13-mers，13bp）处，产生一个开放复合物（open complex)，这个过程需要 ATP。用专一性识别单链 DNA 的内切核酸酶处理大肠杆菌 DNA 和 DnaA 混合物时，该 DNA

的起始区被切开，证实起始区处于熔链状态。

DnaB 是大肠杆菌解旋酶（helicase），是同亚基六聚体，它像夹钳一样在开放复合物处紧紧套住每条单链。DnaB 蛋白在开放复合物上位于 DnaA 和 DnaC 之间，它对 DNA 的结合也需要 ATP。DnaC 蛋白由 *dnaC* 基因编码，其功能是将 DnaB 蛋白护送到 DnaA 处，产生引发前体（图 4-22）。

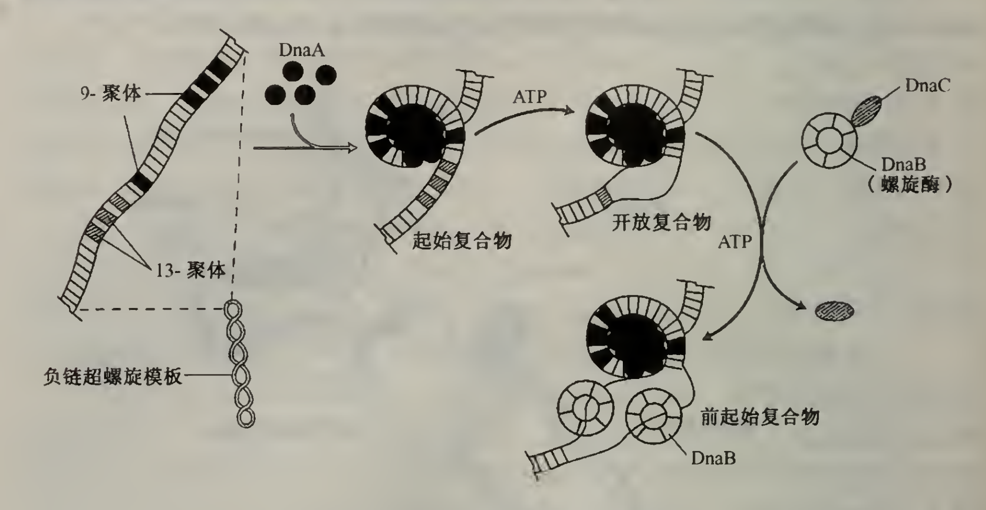


图 4-22　大肠杆菌 oriC 处复制起始模型

解旋酶是一类能利用 ATP 水解的能量沿 DNA 双螺旋移动并且使两条链解开的酶。被解开的单链接着被单链结合蛋白（SSB）保护而阻止退火。当将 *dnaB* 的温度敏感突变株转移到非允许温度培养时，解螺旋停止，结果由于缺少单链模板而使 DNA 的合成也立即停止。像 DnaB 这样的解旋酶结合于单链 DNA 上并沿着该单链以 5′→3′方向移动，同时解开两互补链之间的氢键。由于 DnaB 蛋白像夹子一样套在 DNA 单链上，因此它不容易从 DNA 上脱落。

大肠杆菌中 DNA 引发酶为 DNA 的复制合成引物。大肠杆菌的 *dnaG* 基因编码引发酶，DnaG 蛋白首先与 DnaB 结合的单链 DNA 结合，引发前体与 DnaG 结合后称为引发体，其中的 DnaG 以单链 DNA 为模板合成短的 RNA 引物。

DNA 聚合酶 I 在 DNA 复制期间填补缺口，DNA 聚合酶 II 在 SOS 应答中起作用，DNA 聚合酶 III 催化链延伸。DNA 聚合酶全酶是由 10 个亚基组成的高度复杂的巨大的蛋白质：α 亚基有催化活性；ε 亚基有 3′→5′外切活性，在 DNA 复制中除去错配的核苷酸；θ 亚基的功能尚不清楚；其余亚基的功能是将只能合成 10～15 核苷酸的核心酶（distributive enzyme）转变成能合成 5×10^5 个核苷酸而不脱离模板的酶（processive enzyme），这一活性对于前导链和后随链的合成都是必要的，其关键在于 β 亚基二聚体夹子在复制叉处的 3′端与保持催化活性的核心酶发生联合。与 β 亚基联合后的核心酶移动时，β 亚基二聚体就像戒指套住一个

细线一样沿着 DNA 自由滑动。以这种方式核心酶在复制叉处具有最大的合成长链 DNA 的活性（processive enzyme activity）。其余的六个亚基中，γ、δ、δ′、χ 和 ψ 形成复合物，称为 γ 复合物。γ 复合物执行两个任务：①将 β 亚基运载到 DNA-RNA 引物复合物上；②DNA 复制完成后，卸载 β 亚基夹子。β 亚基夹子的装载或卸载都需要将夹环打开，目前还不明确 γ 亚基是如何完成这个任务的。τ 亚基的作用是使两个核心酶二聚化，这对保证前导链和后随链的合成同时进行是必需的。

当 DNA 聚合酶将脱氧单核苷酸按模板要求逐个加到 3′-OH 末端时，新生 DNA 链开始合成和延伸，延伸的方向是 5′→3′。大肠杆菌环状 DNA 基因组复制从 oriC 开始，进行双向合成。复制中的大肠杆菌染色体在电镜下呈 θ 形状，因此也称为 θ 复制。

2. 猴病毒 SV40 的复制

SV40 DNA 的复制起始于复制原点（图 4-23），病毒编码的 T 抗原或称为 Tag 首先识别复制原点并与之结合（图中第①步），它具有解旋酶活性使 DNA 局部熔链。SV40 复制原点的解链也需要 ATP 和复制蛋白 A（replication protein A, RPA），RPA 是由宿主编码的单链结合蛋白，功能类似于大肠杆菌中的 SSB 蛋白（图中第②步）。一个分子的聚合酶 α（polα）紧密地与引物酶联合，然后结合到每条单链模板上。引物酶合成 RNA 引物，polα 则将引物延伸一小段，在复制原点处以相对的方向形成最早的一段前导链（图中第③步）。polα 的这种活性被复制因子 C（replication factor C, RFC）所激发。然后 PCNA（proliferation cell nuclear antigen）在引物-模板（primer-template）3′末端结合，取代两个前导链模板上的 polα，从而中断前导链的合成（第④步）。接着，polα 结合到正增长着的链的 3′末端，polδ 与 PCNA 的联合增强了聚合酶合成长链的活性，前导链可以连续合成而不再中断（第⑤步）。PCNA 与大肠杆菌中 DNA 聚合酶 III 的 β 亚基高度同源，其氨基酸顺序相似，都能形成戒指环状结构，套住 DNA 细线。但是 β 亚基夹子是二聚体而 PCNA 是三聚体。

随着 Tag 六聚体催化双螺旋 DNA 熔链的进行，Tag 离开复制原点而前进。polα 和 RFC 与后随链模板结合，合成后随链（图中第⑤步）。最后跟原核生物一样，由 topo 酶在链释放和链分离中起作用。SV40 的复制也是以 θ 形式进行。

3. 线粒体 DNA 的复制

无论是原核染色体还是真核染色体，复制子的复制原点都是静态结构：它们是位于双螺旋形式的 DNA 中，在合适的时间用于被识别起始复制的 DNA 顺序。起始需要 DNA 解链并进行双向合成；但线粒体中，其复制原点的排列与其他染色体不同。

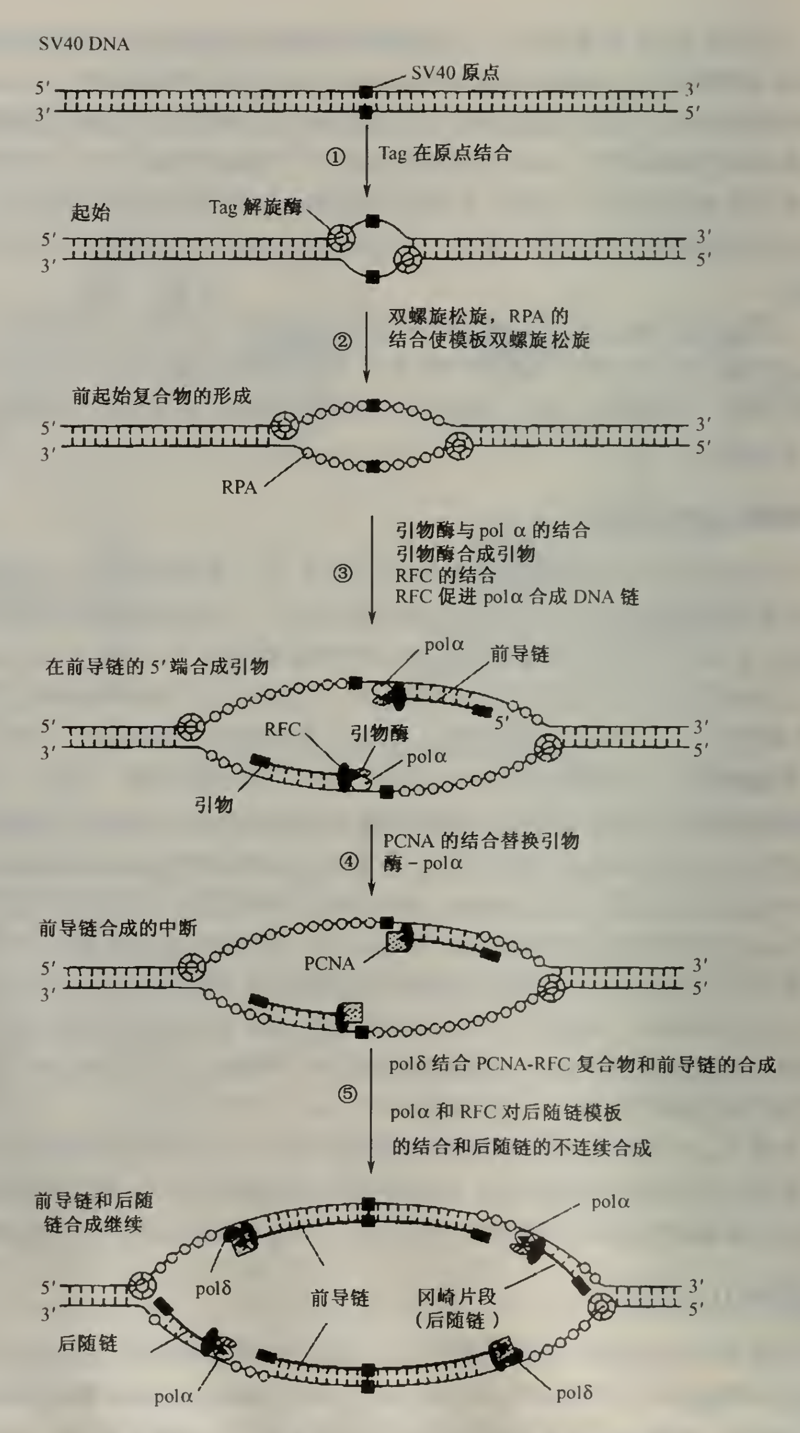


图 4-23　由真核的酶进行的 SV40DNA 体外复制的模型

线粒体 DNA 复制也是从复制原点开始，但起初只复制两条链中的一条链(哺乳动物线粒体的 H 重链)，仅以 H 链为模板先合成一小段（大约是 500～600bp)，替换原来与 H 链互补的 L 轻链（图 4-24)，这种区域的这种状况称为 D 环。有线粒体不止一个 D 环，这是多复制起始区的缘故。

D 环复制中的短链继续扩展，H 链的复制被延伸。当 H 链复制到 2/3 时，L 链上的复制原点被起始，以相反的方向合成 H 链，最后形成两个环状子代分子(图 4-24)。

4. 某些噬菌体的复制

复制产生的结构取决于模板和复制叉之间的关系。关键的特征是究竟是环状 DNA 还是线状 DNA，以及复制叉究竟是同时从两条链合成还是一条链合成。

一些 DNA 分子仅仅复制一条链产生许多拷贝。环状 DNA 分子的一条链上产生一个切口，切口的 3′-OH 末端被 DNA 聚合酶延伸，新合成的链与模板链互补结合替换原来的亲代链。这种结构叫滚环（rolling circle)。

滚环复制在细菌噬菌体中是普遍的，一个基因组单位可以由滚环复制产生的尾巴上切割产生，切割的单链被包装成噬菌体。滚环复制最早是在噬菌体 *ΦX*174 复制研究中发现的。实验发现噬菌体 *ΦX*174 初始合成的单链产物比单位长度要长得多，而且和亲代分子是连在一起的。

*ΦX*174 是单链环状 DNA 噬菌体，该单链称为“+”链；在复制早期，以“+”链为模板合成互补链，称为“-”链，连接后形成双链闭合环状 DNA（称为复制型，replication form，RF)。由病毒编码的蛋白质 A 结合于 RF 型 DNA 的复制原点，在“+”链上产生切口；原来的 5′端可以附着在膜上，留下 3′-OH 作为引物，以负链为模板连续合成比原来的长度长几倍的正链。滚环复制中，当双链 DNA 以“-”链为模板合成时，“+”链以上述方式“滚出”。在 *ΦX*174 中，“滚出”的“+”链按基因组单位被切割、环化和包装。而在其他一些噬菌体如 λ 噬菌体，感染早期其线状双链 DNA 环化，θ 复制一次后进行滚环复制。复制原点产生切口，以切口的 3′-OH 为引物、以未被切开的链为模板进行合成，另一条链被“滚出”，再以“滚出”的链为模板合成双链 DNA；然后由专一的蛋白识别专一位点切割产生基因组大小的片段并进行包装。

5. 线状 DNA 复制问题

前边我们所讨论到的主要是环状 DNA（如大肠杆菌、线粒体及有些病毒)和长的线状 DNA 分子中一个片段（如真核染色体）的复制。但是线状 DNA 包括真核 DNA，它们末端的复制是如何进行的？这是下边我们要讨论的问题。

众所周知，核酸聚合酶（包括 DNA 聚合酶和 RNA 聚合酶）都是从 5′→3′方向合成 DNA。DNA 聚合酶从一条链的末端开始以 5′→3′进行聚合反应，等合成

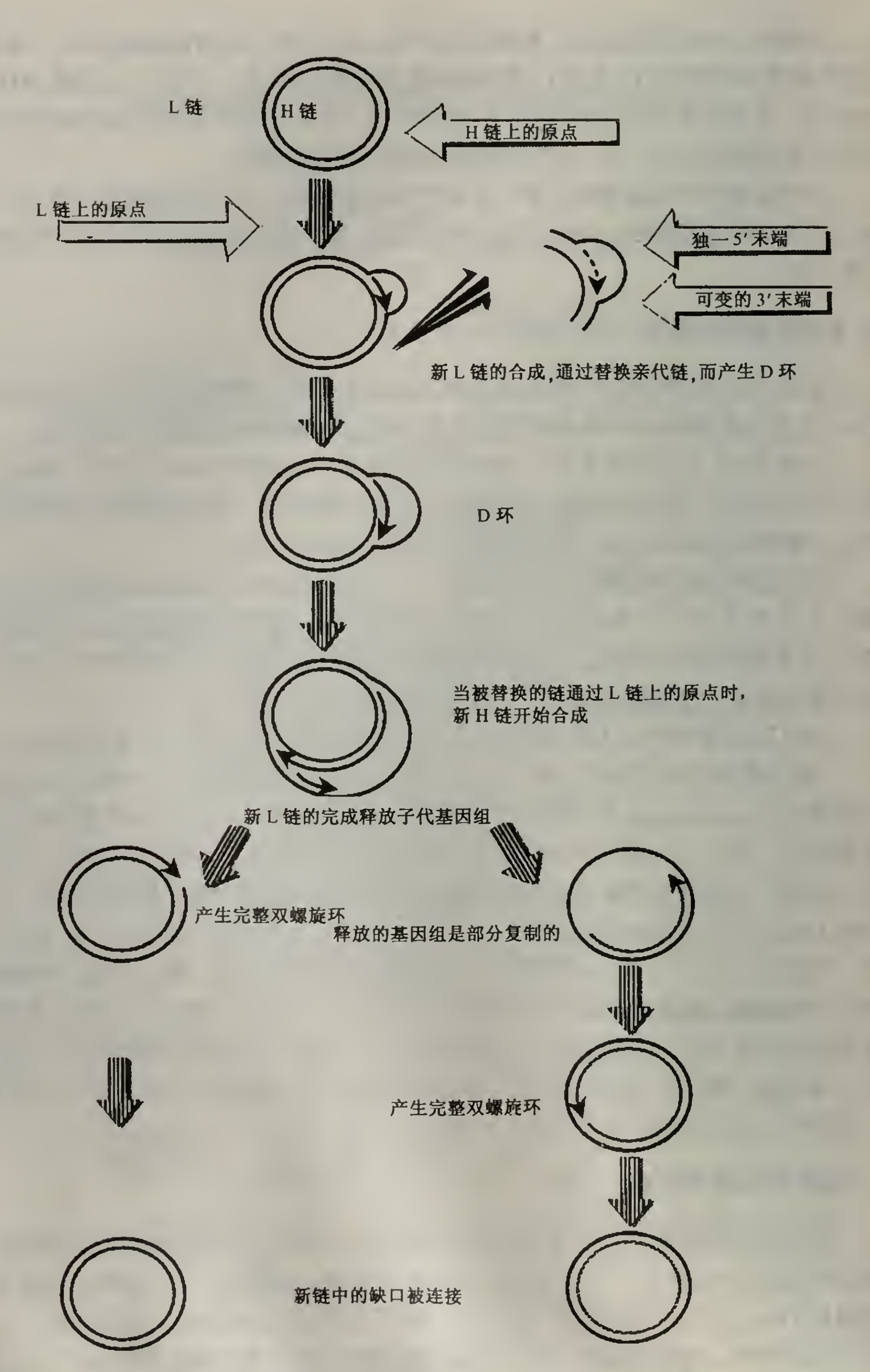


图 4-24　动物线粒体 DNA 复制中维持一个开放 D-环，每条链具有各自的复制原点

完成后聚合酶可能从 DNA 上掉下来，那么另一条链又怎么起始合成呢？因此必须有一特别的机制用于线状 DNA 的复制。可以设想几类模型来适应线状 DNA 末端复制的需要。

1) 将线状 DNA 转换成环状或多分子 DNA 来避开这个末端复制问题，λ 噬菌体和 T4 噬菌体 DNA 的复制就是采取这种方式。

2) 线状 DNA 可以形成一个发夹结构，这样一来实际上就不存在末端，草履虫属的线状线粒体 DNA 中就牵涉到连锁的形式。

3) 线状 DNA 的末端是可变的，其末端存在许多短的重复序列，这些重复单位的数目可以改变，可以增加或减少几个重复单位，从而使复制不必从准确的末端开始。真核染色体可能采取这种方式。

4) 最直接的解决办法是蛋白质的干预，使从真正的末端起始复制成为可能。好几种病毒的线状核酸的 5′末端碱基共价结合有蛋白质，如腺病毒 DNA、ϕ29 DNA 和脊髓灰质炎病毒 RNA。

以腺病毒为例，腺病毒 DNA 复制实际上是从两个真正的末端起始，利用链替换机制，在一端开始合成一条链替换原来的与其模板配对的那条链。当复制叉到达另一端时，被替换的那条旧链作为一个自由单链被释放出来；这种自由单链末端之间有短的碱基互补顺序，可形成锅柄式的结构，双链区作为复制原点可独立地复制该自由单链（图 4-25）。

有几种病毒使用以下这种复制机制，即有一种蛋白质分别结合在 DNA 分子两个末端。腺病毒中有一种 55kDa 的蛋白质（terminal protein，末端蛋白）通过磷酸二酯键跟丝氨酸的结合而与成熟病毒的 DNA 相连接（图 4-26）。同样的情况也在噬菌体 ϕ29 中发现。在脊髓灰质炎病毒中，是一种只有 22 个氨基酸的 Vpg 蛋白，能通过酪氨酸的羟基与 5′末端碱基相连接。这类连接的蛋白均是病毒编码的。

在腺病毒复制时 DNA 链上结合的蛋白质是 80kDa，比从病毒粒子中分离的这种蛋白质要大。这种 80kDa 蛋白质是 DNA 复制起始相关蛋白，在病毒成熟过程中则被切割成较小的分子（55kDa）留在病毒粒子中。

蛋白质对 DNA 的接触是如何克服线状 DNA 分子复制的起始问题呢？体外实验发现，80kDa 蛋白能与腺病毒感染细胞的抽提物中的 dCTP 结合，同时这种蛋白也与另一 140kDa 的 DNA 聚合酶蛋白紧密结合，这些结果引出了图 4-27 中的起始模型。

此 80kDa 蛋白与 140kDa 的聚合酶形成复合物，该复合物携带着 dCTP 结合到腺病毒 DNA 的末端。dCTP 的自由 3′-OH 末端用以起始在 DNA 聚合酶作用下的延伸，产生一个 5′末端连接有 C 的新链。（这个反应实际上牵涉到 DNA 上末端结合蛋白的替换而不是重新结合，腺病毒 DNA 的 5′末端已被上一轮复制时的末

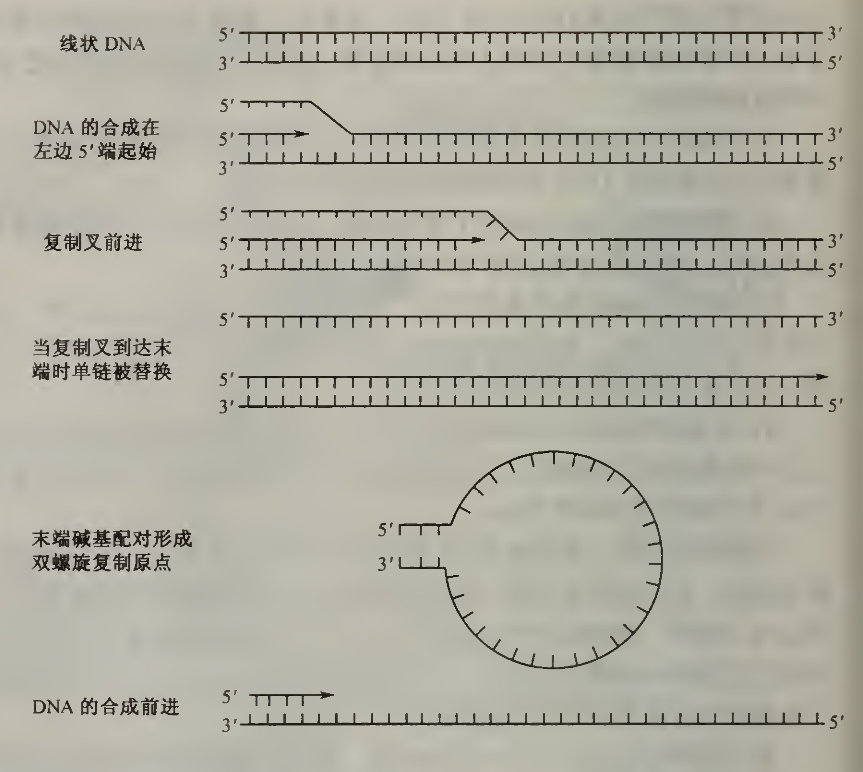


图 4-25　腺病毒 DNA 复制从两端分别起始，并通过链替换前进

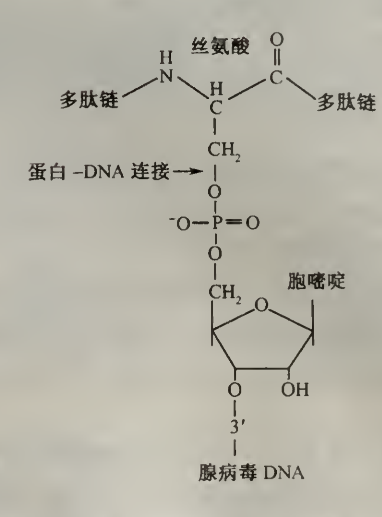


图 4-26　腺病毒 DNA 每个 5′磷酸末端均与 55kDa Ad-结合蛋白中的丝氨酸共价连接

端结合蛋白所结合，在新一轮的复制时，新的末端结合蛋白替换旧的末端结合蛋白。）

80kDa 末端结合蛋白结合到 DNA 末端的 9～18bp 之间的区域，这个核心序列也同样在宿主核 DNA 中发现，位于 17～18bp 之间的侧旁序列对宿主蛋白核因子 1 的识别结合是必须的，这种结合对复制的起始同样重要；因此，起始复合物可能在 DNA 末端上 9～48bp 之间形成，离开末端有一个固定的距离。

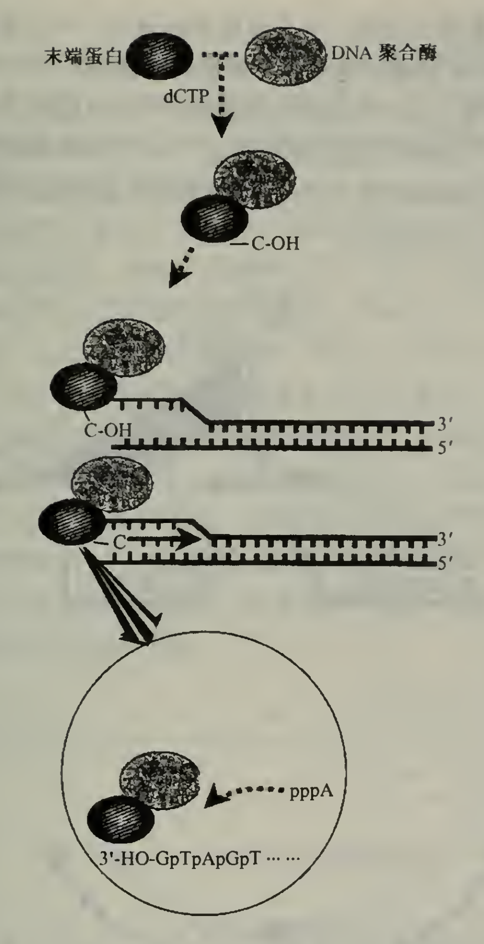


图 4-27　腺病毒末端蛋白结合于 DNA 的 5′末端为新 DNA 链的合成提供一个 C-OH 作引物

四、复制的延伸和终止

链的延伸是在 DNA 聚合酶催化下进行的，它可在引物的 3′-OH 上一个接一个地按模板要求延伸新生链。大肠杆菌 DNA 聚合酶 III 是延伸 DNA 的主要酶。在复制中由 DNA 聚合酶 I 的 5′ →3′的外切核酸酶活性去除引物 RNA，由其聚合活性填补去除引物的空缺。连接酶连接切口，拓扑异构酶则使 DNA 分子分离。

环状双链 DNA 双向复制中存在一个问题：复制在什么地方终止?

一般来说，链的终止不需要特定的信号，也不需要特殊的蛋白质来参与。双向等速复制的环状双链 DNA 分子其复制终点就在离其复制起点 180°的位置上。缺失突变的 λ 噬菌体 DNA 的复制证明了这一点。

大肠杆菌 DNA 复制的终止具有自己的特征。一般从复制原点开始在各复制

染色体一半的位点两复制叉相遇，并且复制在此终止。一个有趣的问题是，由什么东西来保证复制 DNA 正确穿过复制叉相遇的区域呢？

研究发现大肠杆菌上存在引起终止的顺序叫 ter 位点，一个 ter 位点含有 23bp 左右的短序列，在体外可引起复制终止。ter 位点序列只引起一个方向的终止。终止还需 *tus* 基因编码的蛋白质，该蛋白识别 ter 位点一致性序列并与之结合防止复制叉的进程（图 4-28）。

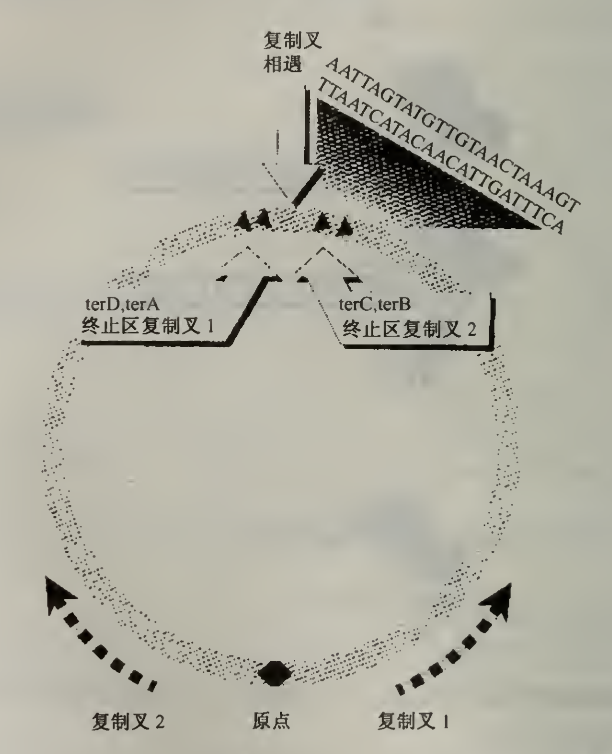


图 4-28　*E. coli* 中复制终止区位于复制叉实际相遇处的两侧

大肠杆菌染色体有两个分别对应于两个复制叉的终止区（terD，terA 和 terC，terB），分别位于相遇位点两侧 100bp 的地方。每一个终止区为一个方向移动的复制叉所特有，而且以这种方式安排每个复制叉将必须通过另一个复制叉的终止区以便达到其敏感的终止区。这种存在两个终止区的安排使即使由于某种原因一个复制叉变慢，另一较快的复制叉也会在其终止区等待慢的复制叉的到达。

第四节　DNA 复制的调控

分子生物学家们并不满足于对 DNA 复制过程的观察和描述，而是希望了解复制过程中各种酶和蛋白质的作用，它们与 DNA 的关系以及复制是怎样受到调

节和控制的。

现以 λ DNA 为例来说明原核生物中 DNA 复制是怎样调节的。

噬菌体 λ 以及其他若干种温和噬菌体是非常令人感兴趣的。因为它们具有进行自主复制和处于溶原状态两种方式，所以这类温和性噬菌体成为我们了解原核生物中 DNA 复制调控过程的良好实验材料。

Jacob 等人首先研究了复制子的调节，发现其中有一个特定的复制起始点可以作为一个整体而被复制。最简单的复制子之一是 λ 噬菌体中占 10% 的一段 DNA，称为 λdv。λdv 是人工制造的环状 DNA（图 4-29）。它包括 O_L、P_L、*cI*、O_R、P_R、*Cro*、*cII*、*ori*、*O*、*P* 等基因。其中 *cII* 不是自我复制所必需的基因，但 O_R、P_R、*Cro*、*O*、*P*、*ori* 基因都是必不可少的。λdv 与转录的控制基因紧挨在一起。这是一个自我复制子，调控是通过 *Cro* 基因产生的 Cro 蛋白结合在 P_RO_R 处，阻遏 Cro 本身的产生，同时也能调节 P、O 两个产物的生成，因为这两个基因也是从 P_RO_R 开始转录的。基因 *O* 和 *P* 的产物是复制的引发蛋白，和基因 *Cro* 是同步转录的。从 P_R 转录一次，不仅产生新的 O 和 P 蛋白使 DNA 复制一次，而且产生更多的 Cro 蛋白，使其浓度增加，作为一个阻遏蛋白阻止了新的 O 和 P 蛋白的产生（图 4-29）。

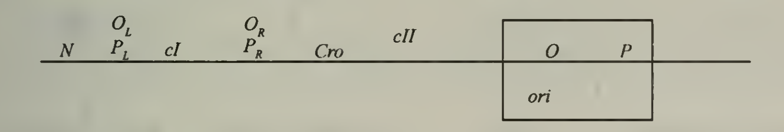


图 4-29　λDNA 的复制控制区

λDNA 带有较多的基因，因此复制的调控比较复杂。现在已经明确 λ DNA 进入大肠杆菌后不久，细菌内就会产生若干种阻遏蛋白。它们是 *cII* 基因和 *Cro* 基因的产物，能作用在 λ DNA 的 O_R 和 O_L 位点上阻止 λ DNA 的转录，阻遏蛋白的多少可以通过调节 O 蛋白的含量来决定 DNA 采取哪一种复制方式。在寄主受到噬菌体 λ 感染的早期 O 蛋白浓度较高，这时 λ DNA 可以双向复制。在后期 O 蛋白浓度较低，则 λ DNA 以单向的滚环方式复制。早期 λ DNA 的复制以 θ 方式进行。这时复制需要 O 和 P 产物，以及寄主的蛋白如 dnaB 蛋白、RNA 聚合酶、引发酶和 DNA 聚合酶 III 等。当 λ DNA 整合到寄主染色体上时，则失去自主复制能力，随寄主 DNA 复制而复制。

通过对噬菌体和病毒的研究，人们现在对原核生物中 DNA 复制的调控过程有了一些了解，当然原核生物 DNA 复制要比病毒和噬菌体复杂得多。原核生物生长和繁殖可以有不同的速率，这取决于培养的条件；但是一旦 DNA 开始复制，DNA 链的延伸速率却是相对恒定的。

真核生物 DNA 的复制又比原核生物复杂得多，至今我们还不清楚真核生物

中DNA复制是怎样控制的；为什么胚胎细胞会迅速地生长而成年细胞生长很慢，有的成年细胞甚至几十年里都不生长；什么因素会使一个正常细胞转变为恶性生长的肿瘤细胞，这些都是生命世界向人类提出的挑战性问题。近年来真核生物DNA的复制和基因的调控正受到越来越多的重视，很可能这方面的研究会为细胞分化、再生、衰老等现象提供说明，也可能为肿瘤的预防和治疗提供有价值的线索，从而为生物世纪的来临铺平道路。

第五章　DNA 损伤修复和基因突变

紫外线引起的 DNA 损伤修复大致通过四种途径：①在损伤部位就地修复——光复活（photoreactivation）；②取代损伤部位——暗修复或切除修复（excision repair）；③越过损伤部位而进行修复——重组修复（recombination repair）；④应急修复——SOS 修复。这四种修复途径如图 5-1 所示。在光复活修复中由光供能，而其他三种修复中由 ATP 水解供能。光复活和切除修复是修复模板，重
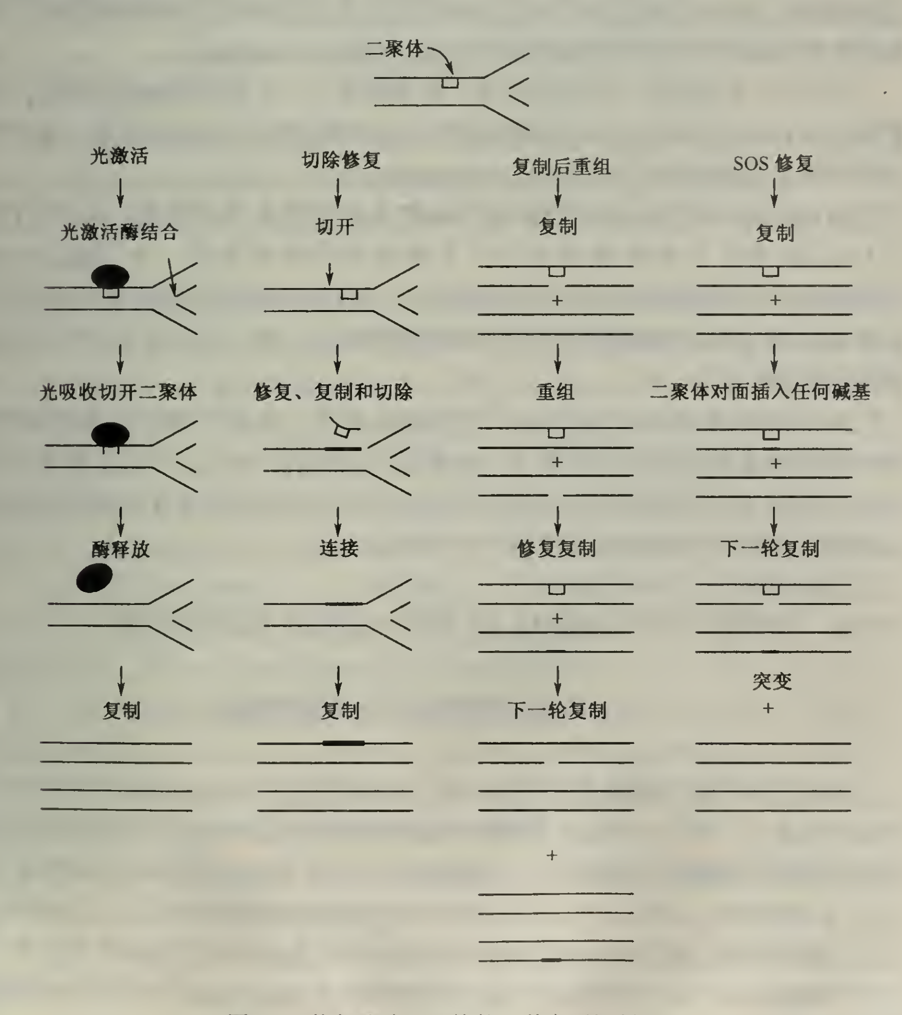


图 5-1　修复嘧啶二聚体的四种主要机制

组修复是形成新模板，SOS修复则为了形成一个不中断的模板来代替有害的模板从而不顾由此带来的危险。SOS修复是唯一导致突变的一种修复系统。

基因突变的研究很早就和DNA损伤以及损伤的修复的研究相伴随。诱变的研究从辐射的诱变作用开始。说明辐射的诱变机制的最早的理论是所谓的靶子学说，它认为辐射的杀伤和诱变作用都是辐射对于细胞中的某些靶子击中的结果。

光复活作用的发现使人们认识到紫外线所造成的DNA损伤并不就是突变本身。这些损伤无非是前突变，前突变通过某些代谢过程才成为突变或导致死亡。以后又陆续发现其他的DNA损伤修复作用，这些修复作用的研究都有助于对基因突变机制的了解。

对于某些诱变剂（如5-溴脱氧尿嘧啶、亚硝酸等）所造成的前突变来说，前突变通过DNA复制而形成突变。这些前突变损伤起着引起编码错误的作用。这种诱变作用可以称为错误编码诱变（miscoding mutagenesis）。

对于另一些诱变剂（例如紫外线、丝裂霉素C）所造成的前突变来说，前突变通过SOS反应的复杂过程而形成突变。这些前突变不起编码作用，这种诱变作用可称为非编码诱变（noncoding mutagenesis）。

20世纪40年代以前对于突变的研究几乎仅限于真核生物。40年代中期M.Demerec等在大肠杆菌中进行了系统的基因突变研究。20世纪50年代S.Benzer进行了噬菌体的基因突变的研究。在DNA双螺旋结构理论的推动下，E. Freese在1969年提出基因突变的碱基置换理论，第一次在分子水平上说明了基因突变的机制。

20世纪70年代以来由于分子遗传学的发展，很多问题已初步得到阐明。DNA损伤修复和基因突变的关系，特别是应急修复（SOS）方面正在进行广泛深入的研究，这方面的研究不但将加深我们对于DNA损伤修复和突变机制的认识，而且可能对于癌变的认识也会有所帮助。

第一节　避免差错的DNA损伤修复

一、光复活作用和切除修复

1927年发现射线诱发基因突变，第二年就发现了260nm波长的紫外线具有最强的杀菌力，而这一波长正是嘌呤和嘧啶吸收最强的波长，因而认识到DNA分子受到损伤是导致细菌死亡的主要原因。1944年通过细菌转化研究证明了DNA是遗传物质，因而进一步认识到遗传物质的损伤是导致死亡的主要原因。

1944年发现经紫外线照射后的细菌可以由可见光的照射而提高存活率，说明受到照射的DNA可以被修复。1950年病毒学家R. Dulbeco发现被紫外线照射的噬菌体所感染的细菌用强的可见光照射，存活的噬菌体数便大为增加，这种间

接的 DNA 损伤修复效应暗示具有 DNA 损伤修复功能的酶可以被可见光所激活。体外实验结果接着证实了这一结论。

R.B.Setlow 在 1954 年首先报道了细菌的 DNA 经紫外线照射后出现的损伤主要是形成嘧啶-嘧啶（环丁烷）二聚体，尤其以形成胸腺嘧啶二聚体为主。光复活酶则作用于二聚体而使它复原为嘧啶。

同年 R.B.Setlow 又发现大肠杆菌的突变型 uvrA、uvrB 和 uvrC 对于紫外线的杀伤作用格外敏感。为了说明原因，他把^{3}H-胸腺嘧啶标记的 uvr 突变型和野生型细菌用紫外线照射，培养一段时间后分别测定它们的胸腺嘧啶二聚体的量，发现 uvr 突变型和野生型细菌的培养物的总二聚体的量并无区别，然而突变型的培养液中的二聚体的量显然比野生型的少。这说明细菌除了可以通过光复活作用而使二聚体复原以外，还可以通过把二聚体切除来修复受到损伤的 DNA。随后证明了 *uvrB*、*uvrC* 编码内切核酸酶。另外一些实验证明了被切除的部分必须通过 DNA 合成来修补。

了解得最清楚的切除修复的例子是大肠杆菌的 uvrABC 系统。图 5-2 阐明了 uvrABC 系统是如何对损伤的 DNA 进行修复的。首先两分子的 Uvr A 和一个分子的 Uvr B 形成复合物，然后结合到 DNA 上，复合物的形成和与 DNA 的结合都需要 ATP。

UvrA-Uvr B 复合物似乎是首先与无损伤的 DNA 结合并沿着 DNA 双螺旋转移直到一个复合物可以识别的变弯曲的部位；沿双螺旋的转移过程同样需要 ATP。与 UvrA-UvrB 复合物结合的 DNA 损伤区在 ATP 作用下，构型变化使 DNA 骨架产生弯曲。当 UvrA 二聚体解离后，具有内切核酸酶活性的 UvrC 蛋白结合到损伤位点。UvrC 和弯曲的 DNA 的相互作用被认为是开放 DNA 中的空间允许酶的残基进入其靶位点的过程（图 5-3)。切割的准确位点已经测定，Uvr C 切割嘧啶二聚体两侧的两个磷酸二酯键：一个位于 5′端距损伤处 8 个核苷酸处，另一个位于 3′端距损伤处 4～5 个核苷酸处。当损伤链的两个位点被切割之后，这个被切下的片段和 UvrBC 复合物一起被 DNA 解旋酶移去并且被降解，链中的缺口通过 DNA 聚合酶 I 和连接酶的联合作用来修复。

UvrABC 也能切除丝裂霉素 C、补骨脂素和亚硝酸所造成的 DNA 损伤。对于真核生物来说，紫外线等对遗传物质造成的损伤和原核生物没有多大区别，不过由于真核生物中的染色体蛋白质与 DNA 构成了复杂结构，所以修复过程也比原核更为复杂，特别是涉及多种酶的切除修复更是如此。

20 世纪 50 年代发现了酵母的光复活现象，后来发现了酵母有关光复活的基因，70 年代又发现了切除修复现象。

着色性干皮症（一种人的遗传性疾病）是典型的由于切除二聚体的能力的缺损而带来的疾病，而且主要是由于内切核酸酶功能丧失的结果。通过测定异核体细胞的修复功能，可以把干皮症基因区分为 7 个互补群，这说明真核生物的切除

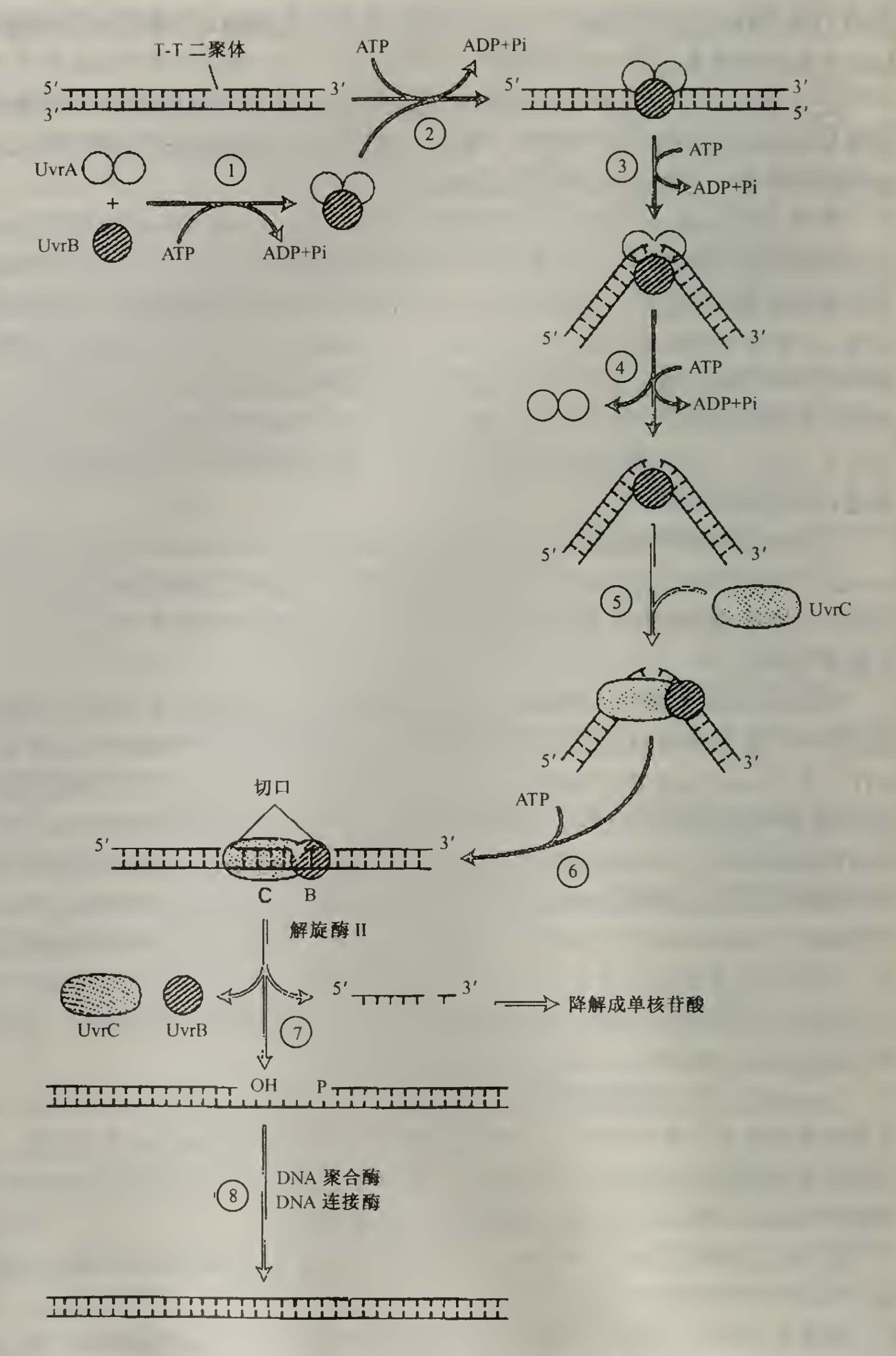


图 5-2　由 *E. coli* UvrABC 进行的 DNA 切除修复的机制

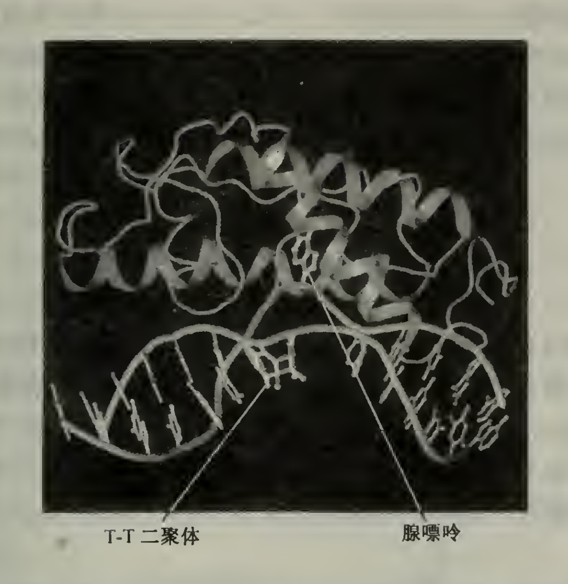


图 5-3　根据 X 射线结晶学分析噬菌体 T4 内切核酸酶 V 和含有 T-T 二聚体的 13bp DNA 片段形成的复合物模型

修复系统可能比原核生物更为复杂。

从切除的机理来看，整个过程也显得更为复杂。因为在核小体部分的二聚体 DNA 损伤切除前，核小体结构必须先行解开而使损伤部分暴露，损伤部分经切除并进行修复合成以后又重新装配成核小体。

二、重组修复

A.J.Clark 等在 1965 年发现了大肠杆菌的重组缺陷型 recA，并且发现 recA 对于紫外线的杀菌作用格外敏感。不过 recA 和 uvr 突变型不同，它和野生型一样能进行光复活修复和切除修复，说明细菌除了光复活修复和切除修复以外还有另一些修复功能，recA 细菌由于这种修复功能的丧失而使它对紫外线格外敏感。

把 uvr 突变型细菌用紫外线照射并在^{3}H-胸腺嘧啶核苷的培养基中培养 10min，提取 DNA 并在碱性蔗糖梯度溶液中离心，发现它比未经照射的细菌 DNA 沉降得慢，但经 70min 培养后两者 DNA 的沉降速度相同。对于上述现象的解释是 DNA 复制过程中新合成的单链在二聚体的位置留下空隙，因此紫外线照射后最初合成的单链是不连续的，在进一步培养过程中由于发生了重组而出现连续的子链，这就使得它的沉降趋于正常。

实验结果还说明在 DNA 复制过程中二聚体会随机地分布到新合成的子链中，与此同时切除修复也继续进行。可见重组修复并不能除去二聚体，它不过能

使细菌在依赖切除修复来消除所有的二聚体之前便能出现正常复制的细胞。

真核生物中也发现有类似的现象。可是真核生物有多个复制起点，每个复制单位进行双向复制，因而在正常复制过程中 DNA 子链本来就是一段一段地连接而成的，这为实验结果的解释带来了困难，因此关于真核生物中的重组修复还有争论。

三、N-糖苷酶和 DNA 损伤修复

DNA 聚合酶 I 和聚合酶 III 在 DNA 复制过程中对于 dTTP 和 dUTP 的分辨能力很差，因此少量 dUTP 可能掺入到 DNA 分子中去。大肠杆菌虽然具有 dUTPase，它能使 dUTP 变为 dUMP，但是由于它的效率很低，所以仍然不能完全避免 dUTP 的掺入。此外胞嘧啶经脱氨基作用后可以形成尿嘧啶。对于 DNA 分子上出现的尿嘧啶，细胞中的尿嘧啶 N-糖苷酶（uracil N-glycosylase）可以把它除去。

N-糖苷酶有多种。大肠杆菌的 3-甲基腺嘌呤-DNA 糖苷酶 II（3-methyladenie-DNAglycosylase II）是一种诱导酶，它的专一性不强，对于 7-甲基鸟嘌呤、3-甲基鸟嘌呤和 7-甲基腺嘌呤也有作用。它由基因 *alkA* 所编码。这种诱导不依赖于 SOS 反应。

另外一种修复烷基化作用所造成的 DNA 损伤的诱导酶是 O-6-烷基鸟嘌呤-DNA-DNA-烷基转移酶，这也是一种与 SOS 反应无关的诱导作用产生的修复功能。

此外还有一种糖苷酶 3-甲基腺嘌呤-DNA 糖苷酶 I，它专门作用于 3-甲基腺嘌呤，为 *tag* 基因所编码。它在大肠杆菌中不需诱导便能行使修复功能。

过去认为胸腺嘧啶二聚体的修复与糖基化酶无关，近年来的研究结果说明在二聚体切除修复中同样涉及 N-糖苷酶和 Ap 内切核酸酶（apurinic endonuclease, Ap endonuclease）。

修复作用首先通过 N-糖苷酶的作用把 5′端的胸腺嘧啶和五碳糖之间的 N-糖苷键打开，然后在 Ap 内切核酸酶的作用下把无嘌呤位置的 3′端的磷酸二酯键打开，进一步又分解不带有碱基的磷酸脱氧核糖，然后在聚合酶 I 和连接酶的作用下完成修复作用。

四、校读作用

在 DNA 复制过程中偶然地可能把一个不正常的核苷酸加到 3′-OH。实际上大肠杆菌 DNA 聚合酶 III 的 α 亚基在体外复制系统的错误掺入率为 $1/10^4$，大肠杆菌一个基因的大小平均为 10^3bp，其复制时的错误掺入率为 1/10，然而细菌细

胞检测出的突变频率要低得多，聚合 10^9 个核苷酸只有一个错误，也就是说每个基因每世代的突变频率为 $10^{-5} \sim 10^{-6}$。

这种在体内复制准确度的增加主要是由于大肠杆菌 DNA 聚合酶的校读功能。大肠杆菌 DNA 聚合酶 I 的 3′→5′外切活性可将正在合成中的引物－模板复合物的新链 3′末端的错配碱基切除。DNA 聚合酶 III 核心酶的 ε 亚基也有这种功能。当 DNA 合成期间一个不正确的碱基被掺入，聚合酶的聚合活性暂停，将正在增长的链的 3′末端转移到其外切活性位点，错配的碱基在此被除去，然后该 3′末端再被转移到聚合酶活性位点，在此正确的碱基被掺入（图 5-4）。校读功能是所有 DNA 聚合酶的一个性质。动物细胞中 DNA 聚合酶 δ 和 ε 也具有校读活性，但 DNA 聚合酶 α 无此校读活性，这些功能似乎是所有细胞为避免过多的遗传损伤而必不可少的。

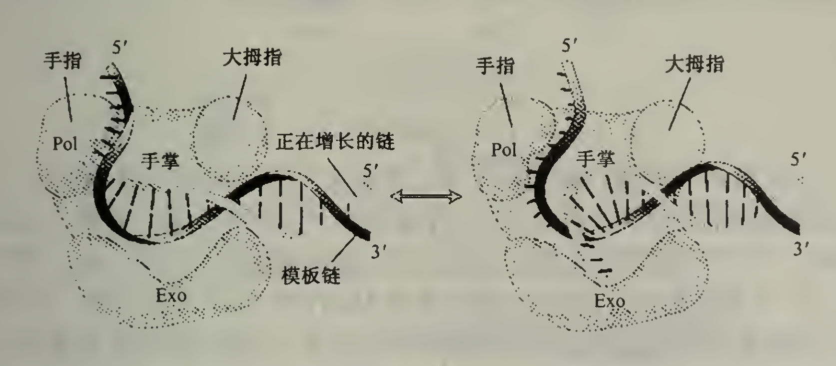


图 5-4　DNA 聚合酶校读功能示意模型

所有的 DNA 聚合酶都有类似的三维结构，它们装配成一个半张开的右手形状。手指和模板的单链片段结合，聚合酶在手指与手掌交界处发挥作用，当一正确碱基掺入到增长链的 3′末端，3′末端留在聚合活性位点。当不正确的碱基掺入时，聚合活性停止，3′末端被转移到 3′-5′外切活性位点，错配碱基被除去后又回到聚合活性位点继续进行链的延伸。

噬菌体 T4 的一个增变基因便是编码聚合酶 I 的基因（基因 *43*）发生突变的结果。

大肠杆菌另一个增变基因 *dam* 编码腺嘌呤甲基化酶（DNA adenine methylase），这一事实说明腺嘌呤甲基化酶是其校读系统的一个成员。关于 dam^+ 基因在校读中所起的作用以及 *dam* 是增变基因的解释是这样的：DNA 的甲基化是限制修饰功能的一部分，甲基化发生在 DNA 合成以后，细胞中的 DNA 都是已经在前一轮复制中被甲基化了的。新合成的 DNA 逐步被甲基化，越是靠近复制叉处甲基化程度越低，错配修复系统能识别甲基化和未甲基化的 DNA。不正常的核苷酸不能和正常的核苷酸配对，因此被聚合酶 I 识别并被切除。可是对于某些不正常的核苷酸来说，聚合酶不能把它和正常的核苷酸区别开来，在这种情况下聚合酶 I 就有可能把正常的核苷酸切除而造成校读的错误。错配修复系统能识别

甲基化和未甲基化的 DNA，这就保证了切除的核苷酸是新合成的，即不是正常的核苷酸。由于这一缘故，失去甲基化能力的细菌更容易发生突变，这就是 *dam* 是增变基因的原因。

细菌和真核细胞都具有错配修复系统，可以识别并修复除 C-C 以外的所有单碱基错配，同样也可以修复一个小的插入和缺失。大肠杆菌中甲基化指导的错配修复系统常被归类为 Mut HLS 系统。

在大肠杆菌 DNA 中，GATC 短序列中的 A 在第 6 位被甲基化，在合成过程中 DNA 聚合酶将一个未甲基化的 A 掺入到 DNA 中，新复制的 DNA 中只有亲代链中的 A 是被甲基化的。子代链中 GATC 序列中的 A 到几分钟后被专一性酶即 Dam 甲基转移酶（Dam methyltransferase）甲基化，在后随期中，新复制的 DNA 含有半甲基化的 GATC 序列：

$$\begin{array}{l}\qquad\qquad\ \ CH_3\\ \qquad\qquad\quad |\\ 5'—G—A—T—C—3'\ \ \text{亲代链}\\ 3'—C—T—A—G—5'\ \ \text{子代链}\end{array}$$

大肠杆菌的一种蛋白质 MutH 专一性地与半甲基化的子代链结合。如果 DNA 复制期间发生错误从而引起在 GATC 附近产生错配，另一种蛋白 MutS 则结合到这个不正常配对的片段上（图 5-5）。MutS 的结合会引起 MutL 的结合，MutL 是一个连接蛋白，它可以与附近具有 MutH 的 MutS 蛋白连接。这种交叉连接可以激活 MutH 潜在的内切核酸酶活性从而专一性地切割未被甲基化的子代链，随之这个含有错配碱基的子代链会被切除而正确的 DNA 序列则取而代之。

缺乏 MutH、MutS 或 MutL 蛋白的大肠杆菌的自发突变频率比野生型要高。当然不能合成 Dam 甲基转移酶的菌株的自发突变率也较高。

真核生物同样有 DNA 甲基化酶，其甲基化曾被认为与 DNA 的复制和重组等过程有关。真核生物可能具备比原核生物更为复杂的校读系统。

五、交联的修复

8-氧甲基补骨脂（8-methoxypsorolen）能嵌入 DNA 分子中，在 360nm 光照下很容易和邻近的两个碱基发生反应使它们交联，交联发生在两条 DNA 单链之间。丝裂霉素 C 也有这样的作用。交联会阻碍 DNA 复制，这也是一种具有致死效应的 DNA 损伤。

对交联的修复目前尚不清楚，一种模型认为交联的修复既包括一条单链上一部分核苷酸的切除，也包括重组。

真核生物必然也有交联修复功能。因为离体培养的 Fanconi 氏贫血症患者的细胞对补骨脂的杀伤作用格外敏感，由此说明它的交联修复系统有缺陷，很明

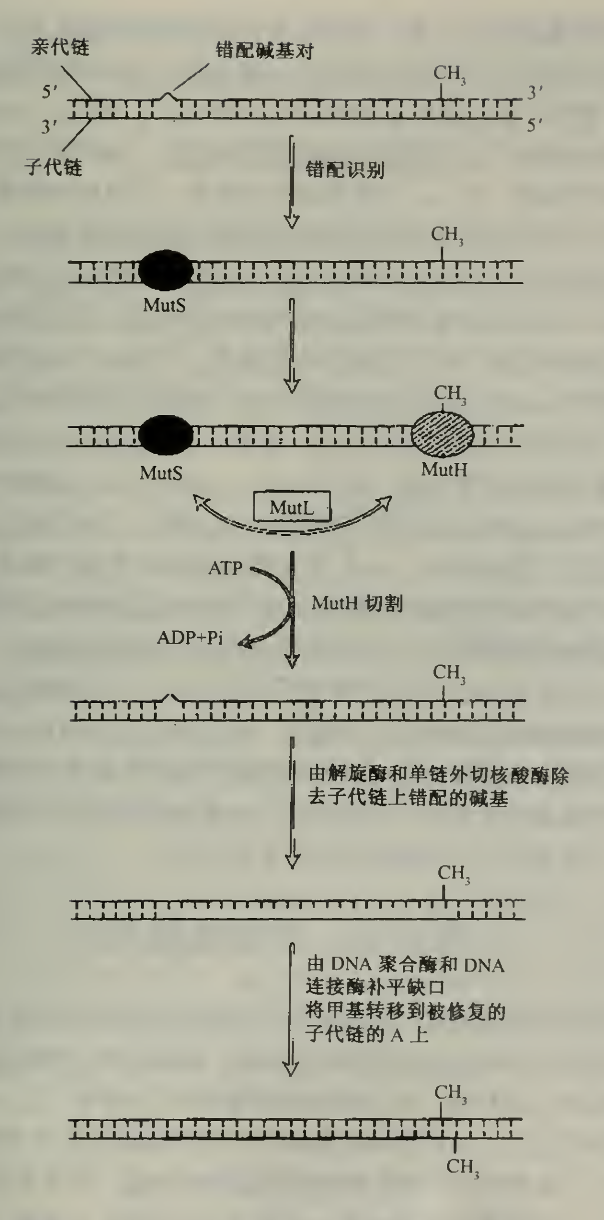


图 5-5 通过 *E.coli* MutHLS 系统错配修复的模型

显，正常人是具有交联修复功能的。

第二节 避免差错的 DNA 损伤修复和基因突变

光复活作用是使嘧啶二聚体分解而恢复 DNA 分子的原来结构，这一简单的

过程并不会带来基因突变，是一种避免差错的DNA损伤修复作用。

切除修复过程是否会带来基因突变，这可以从uvr突变型的诱变行为中找到答案。uvr突变型对紫外线的杀伤作用格外敏感。E. Witkin 1969年报道了紫外线对uvr突变型和野生型的杀伤和诱变效应的比较，结果野生型对紫外线的杀伤和诱变作用都不敏感，但uvr突变型却格外敏感。这说明切除修复过程中并不带来基因突变，相反，保留着嘧啶二聚体的细菌却容易发生突变。这一事实暗示了依赖于recA的修复过程可能会带来突变，因为通过实验已经确定recA突变型能进行切除修复，可是却对紫外线的杀伤作用格外敏感。而正好和uvr突变型情况相反，$recA^-$细菌对紫外线的诱变作用不敏感，而$recA^+$则是敏感的。这一事实说明了在依赖于recA的修复过程的进行中似乎带来了基因突变。可是从1971年开始得到的一些实验结果却否定了这一论断。①*recB*和*recC*突变型对紫外线的诱变作用的敏感性接近野生型。已知*recA*、*recB*、*recC*是属于同一重组途径的三个基因，所以*recB*和*recC*似乎应该和*recA*相似，可是事实上它们却和野生型相似。因此对此只能理解为：recA突变型之所以对于紫外线的诱变作用不敏感不是由于它是重组缺陷型的，也就是说并不是由重组修复过程带来基因突变，造成基因突变的原因可能是与recA功能有关的另一种修复功能。②自从在大肠杆菌中发现了可以通过recF途径实现重组过程以后，就应考虑紫外线也可以通过recF的重组修复而诱发突变这一可能性。可是事实上$recA^+$ $recB^-$ $recF^-$细菌对紫外线的诱变作用是敏感的，这一结果同样说明重组修复并不导致基因突变。*recA*基因在紫外线诱变中的作用并不是由于它在重组中的作用，而是由于它在其他方面的作用。这就是我们下节要讲的应急修复功能。

第三节　应急修复反应

以上两节所讨论的DNA损伤修复功能都是不需诱导的功能。早在20世纪50年代（1953年）Weigle发现如果用经紫外线照射的λ噬菌体感染经低剂量的紫外线照射过的大肠杆菌，存活的噬菌体数便大为增加（这一效应称为Weigle复活效应或W-效应），在存活的噬菌体中出现较多的突变型（Weigle诱变效应或W-诱变效应）。如果感染的是未经紫外线照射的细菌，那么存活噬菌体和突变型的量都较少，可见这些效应是经紫外线诱导后出现的。这些反应就是所谓的SOS反应（应急修复反应），SOS反应除上述两种效应外，还有其他效应，包括λ溶源菌释放噬菌体、DNA损伤修复、细胞分裂的抑制等。

这一诱导现象和大肠杆菌的诱导酶诱导合成现象类似。如果用经紫外线照射的λ噬菌体感染含有氯霉素的培养液中的细菌，那么W-效应就不出现，这说明诱导将会伴随着新的蛋白质的合成。在β-半乳糖苷酶的诱导合成中已经被诱导的大肠杆菌如果不再继续接触诱导物，那么酶的活性又会逐渐下降到未经诱导的水

平。W-效应也是一样，经紫外线诱导后逐渐出现，隔了几十分钟后又逐渐消失。这些现象都说明 SOS 反应类似于诱导酶的诱导合成作用。

一、*recA* 和 *lexA* 基因

recA 和 lexA 突变型都不呈现 W-效应，说明这两个基因都是 W-效应所必需的，它们分别处在大肠杆菌遗传图的 58min 和 91min 位置上。*recA* 基因是重组基因，recA 突变型也称为 recA（Del），它既失去了接合重组能力，又失去了 W-效应，而且对紫外线格外敏感。lexA（In）也失去了 W-效应并对紫外线敏感，可是它的接合重组能力还是正常的。可见 W-效应和接合重组并没有直接关系，recA 蛋白则在两方面都起作用。

$recA^-/recA^+$ 杂合因子呈现 SOS 反应，可见 *recA* 是显性的，说明 *recA* 基因具有正面作用的功能。$lexA^-/lexA^+$ 杂合因子不呈现 SOS 反应，可见 *lexA* 是隐性的，说明 *lexA* 基因起反面作用的功能。

根据 *recA* 有正面功能和 *lexA* 有反面功能这一假设，可以认为 $lexA^+$ 编码一种阻遏蛋白，这一阻遏蛋白阻遏与 SOS 反应有关的一系列基因的表达，$recA^+$ 则编码分解 LexA 蛋白的酶。

RecA 蛋白在 1978 年已被分离出来，在用 λ 溶源性细菌诱导的研究中发现在 λ 的诱导释放的同时会出现一种分子质量为 40 000Da 的蛋白质 X，这一蛋白质就是 RecA 蛋白，它能分解 λ 阻遏蛋白和 LexA 蛋白。1979 年分离鉴定了 LexA 蛋白，分子质量为 24 000Da。

SOS 反应包括一系列反应，这一事实暗示 LexA 蛋白能阻遏一系列的基因。可以设想被诱导激活的 RecA 蛋白能够分解 LexA 蛋白，于是一系列基因的阻遏被消除从而引发一系列的 SOS 反应。通过大量的遗传学研究，估计与 SOS 反应有关的基因可能占大肠杆菌全部基因总数的 0.5%。通过对 SOS 基因的离体研究，已经确定了 LexA 蛋白在 DNA 上的结合部位。RecA 和 LexA 在 SOS 反应中的作用见图 5-6。

二、SOS 反应过程

根据对于 SOS 有关的基因的突变效应的研究以及一些离体研究的结果，可以把 SOS 反应归纳为以下过程：①诱导前。在未经诱导的大肠杆菌细胞中与 SOS 反应有关的全部基因都受到 LexA 蛋白的阻遏，包括 lexA 本身。不过由于各个基因和 LexA 蛋白的亲和力不同以及其他原因，阻遏并不都是完全的。*recA* 基因虽和 LexA 亲和力最大，可是在未经诱导的细胞中 RecA 蛋白仍然会少量合成以供正常遗传重组中的需要；②诱导。紫外线、丝裂霉素 C 等都直接造成

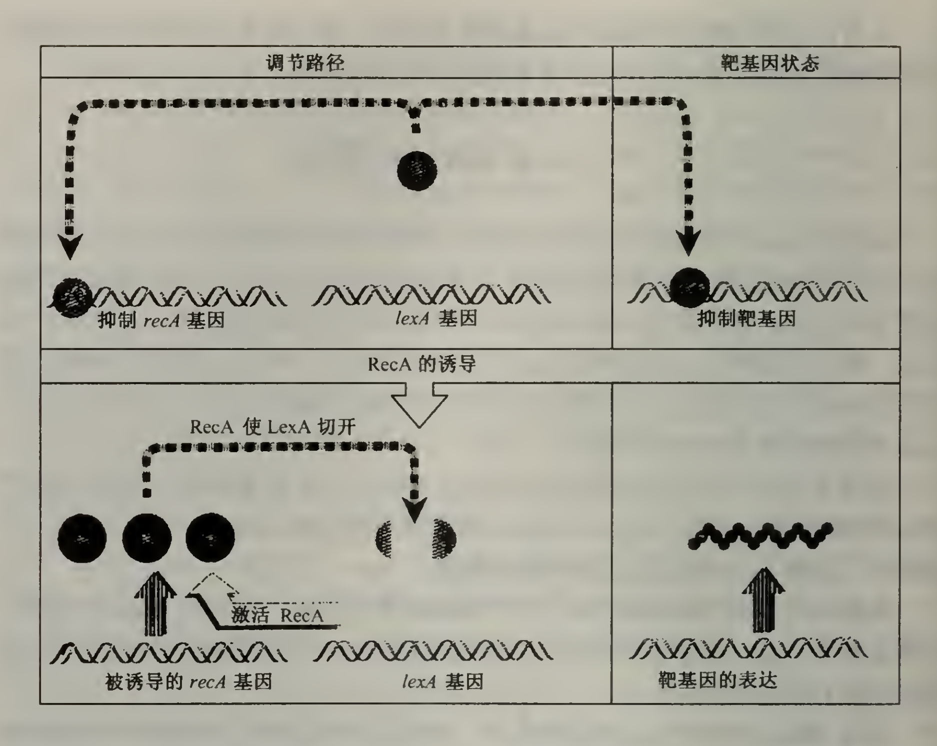


图 5-6 LexA 蛋白抑制包括修复功能在内的许多基因，也包括 *recA* 和 *lexA*。RecA 蛋白导致 LexA 的水解切割而所有这些被抑制的基因被诱导

DNA 损伤，它们都是 SOS 反应诱导物。胸腺嘧啶缺陷型在胸腺嘧啶供应不足时或者一些 DNA 复制突变型（例如 dnaBts）在 42℃ 中也会出现 SOS 反应。可以设想这些因素会促使细胞中产生一种诱导信号，而这一诱导信号才是真正的诱导物。有一种看法认为由于这些处理而出现的 DNA 分子中的单链部分是诱导信号，此外，复制完成后复制叉位置上被分解的 DNA 片段可能也是一种诱导信号。细胞中经常存在的少量 RecA 蛋白结合在这些单链部分后，RecA 蛋白便被激活。激活的 RecA 蛋白作用于 LexA 蛋白和 λ 阻遏蛋白。LexA 蛋白被分解后，SOS 网络中的一系列基因便开始转录，如果这是一个 λ 溶源性细菌，那么 λ 阻遏蛋白被分解后，λ 溶源细菌中的 λ 便开始复制。③诱导状态。由于阻遏消除而首先转录的是 *uvrA*、*sulA* 等基因。*uvr* 基因编码切除修复的内切核酸酶，由于 *uvr* 的转录，切除修复加强，这就有利于 DNA 的损伤修复。如果诱导物剂量较低或者不再存在，那么 DNA 损伤修复之后，诱导信号也就不存在，细胞就恢复正常。*sulA* 是和细菌分裂有关的基因，它编码一种抑制细胞分裂的蛋白质从而抑制细胞分裂。诱导物的继续存在或者在细菌接触高剂量的诱导物的情况下，细胞可以进入充分诱导状态，LexA 蛋白进一步分解使 SOS 网络中的更多基因消除

阻遏，其中包括与诱变作用有关的基因以及作用还不清楚的一些基因。④恢复状态。诱导物不再存在时，诱导信号不再增加，同时由于 DNA 损伤的修复，原有的信号随之消失，于是 RecA 蛋白回到不激活状态，抑制细胞分裂和诱发突变有关的基因又被阻遏，于是整个细胞又恢复正常状态。

三、SOS 反应和细胞分裂

引起 SOS 反应的因素都使 DNA 复制不能正常进行，而 SOS 反应一方面促进细胞的修复功能，另一方面又产生一种抑制细胞分裂的物质——伴随修复的分裂抑制物。一种观点认为 SOS 反应中产生抑制细胞分裂的物质是为了使受到损伤的细胞不产生无 DNA 的细胞，或者使细胞中有更多的进行重组修复的机会。

四、SOS 反应和 DNA 复制

紫外线照射使 DNA 分子上出现嘧啶二聚体，从而阻碍 DNA 复制的进行，但紫外线照射却又能促使两种特殊形式的复制的出现而使 DNA 合成能继续进行：一种是依赖于发生变化的 DNA 聚合酶的 DNA 复制，另一种是不依赖于蛋白质合成而发动的 DNA 复制，即所谓稳定的 DNA 复制。用紫外线照射、^{3}H-胸腺嘧啶标记和加氯霉素处理等方法证明了不依赖于蛋白质合成的 DNA 复制的发动是一种 SOS 反应。

对于这种不依赖于蛋白质合成的 DNA 复制的发动的解释有以下两种：其中一种是 DNA 复制机器（称为复制体 replisome）是由不同的 *dna*$^{+}$ 基因编码的蛋白质所组成的一种不稳定的复合物，这里包括破坏复制体的蛋白质，复制一结束它就使复制体解体，因此每次复制发动前必须进行蛋白质合成，SOS 诱导物则使这种蛋白质消失或破坏，因而复制的发动前便不需蛋白质的合成。另一种可能是紫外线诱导激活 RecA 蛋白使之具有蛋白水解酶活性，由于专一性地作用这一成分从而使复制的发动不需蛋白质的合成。

这些诱导的出现使 DNA 复制在遇到阻遏时仍能够继续进行，它们的生物学意义是显而易见的。

五、SOS 网络中的基因和诱发突变

SOS 的研究从 W-诱变效应开始，当时观察到寄主细胞的 SOS 系统对于所感染的 λ 噬菌体具有修复和诱变作用，后来又发现 SOS 系统也对细菌本身的 DNA 损伤进行修复和诱发突变。

经研究发现除了 *recA* 和 *lexA* 基因以外还有 *umu* 基因参与 SOS 系统。转座

子的插入可以诱发 *umu* 突变，说明 umu^+ 基因具某种正面功能，有促使基因发生突变的作用。umu^+ 基因可以为丝裂霉素 C 所诱导，它们的调控区的结构和 *recA*、*lexA* 以及 *sulA* 基因都相似，这说明它是 SOS 系统中的一员。UmuD 蛋白可被 RecA 蛋白切割而被激活，并且也可以激活倾向差错的修复系统（error-prone repair system）。

六、二聚体和诱发突变

SOS 反应是一种倾向差错的 DNA 损伤修复反应，那么突变发生的位置是否就是受到损伤的位置呢?

紫外线的杀伤作用主要是由于它可以使 DNA 分子上出现胸腺嘧啶环丁烷二聚体，这些二聚体的修复依赖于光复活作用、切除修复和重组修复，这些都是避免差错的修复作用。紫外线还造成另一些损伤，其中一种损伤是嘧啶-嘧啶(6-4)二聚体，大部分是 T-C 和 C-C 二聚体。实验结果发现紫外线诱发的突变都发生在形成二聚体的位置，尤其是嘧啶-嘧啶（6-4）二聚体。

七、无嘌呤位点和基因突变

应用 *ΦX*174DNA 所进行的一系列实验说明，通过 SOS 反应而导致突变的另一种 DNA 损伤是无嘌呤位点（AP），脱嘌呤作用可由烷化剂和低 pH 等因素引起。用完成脱嘌呤反应后的 *ΦX*174DNA 转染脱细胞壁的细胞，加入 DNA 聚合酶 I 和［α-^{32}P］标记的四种单核苷酸测定脱嘌呤作用对 DNA 合成的影响，发现随着脱嘌呤位点的增加，每一模板合成的核苷酸相应减少，说明脱嘌呤与二聚体一样能阻遏 DNA 的复制。另外，回复突变的频率却随无嘌呤位点的增加而增加，说明基因突变是在通过无嘌呤位点的复制过程中发生的。这个过程正是 SOS 反应的一部分，因为脱细胞壁的细胞必须经过紫外线照射才能发生回复突变，而且这一过程依赖于 $recA^+$ 基因，回复突变不发生在不经紫外线照射的脱壁细胞中，而突变是在无嘌呤位点的复制过程中发生的。无嘌呤位点像二聚体一样阻碍复制的进行，由此可见 SOS 反应还包括通过这些 DNA 损伤部位的 DNA 复制。

八、DNA 聚合酶和基因突变

从经嗦啶酮酸诱导的大肠杆菌中曾分离到一种 DNA 聚合酶 I*，这种聚合酶和未经诱导的细菌中的 DNA 聚合酶 I 十分相似。曾用多核苷酸共聚物 poly（dA）、poly（dT）作模板来比较在正常的 DNA 聚合酶或者在 DNA 聚合酶 I* 的作用下 dAMP、dTMP 的互补掺入量和 dGMP 的非互补掺入量，发现对于

互补掺入量来说两者并无明显的差别，可是对于非互补掺入量来说 DNA 聚合酶 I* 显然高出 DNA 聚合酶 I 的 10 倍。

上述结果说明在经 SOS 诱导的细菌中会出现一种新的 DNA 聚合酶，这种聚合酶能够有效地通过 DNA 损伤位置进行 DNA 复制而避免了死亡，可是另一方面却发生了更多的突变。

SOS 诱导使 DNA 聚合酶的外切核酸酶的活力下降，这样即使在二聚体的相对位置上出现了不正常的核苷酸，DNA 复制仍然继续前进。可是同时在这个位置上却增加了出现不正常核苷酸的机会，因此便带来了更多的突变。大肠杆菌聚合酶 II 在反应中也起作用，并且该酶也缺少某些外切核酸酶活性。

第四节　诱变剂、诱变、基因突变和突变体

在讨论基因突变时常常涉及四个名词，即诱变剂、诱变、突变和突变体（mutagen、mutagenesis、mutation 和 mutant）。

诱变剂是指能诱发产生突变的物理和化学因素，例如紫外线照射和一些化学药品如吖啶类等，它们造成 DNA 损伤或改变碱基配对性能，这些物质就叫做诱变剂。

诱变是指用诱变剂处理从而产生突变的过程。

突变是指存在于一个突变体中的 DNA 结构上的任何改变，这种改变能产生一个突变表型。突变常常是改变了 DNA 的碱基顺序，绝大多数是碱基的替换、增加、重排、减少等。

突变体是指某个有机体的遗传状态，正常的个体称为野生型，那么某一表型特征与野生型不同的个体则称为突变体，例如大肠杆菌有利用半乳糖作碳源的性质，它的野生型状态是 Gal^+，Gal^-则不能利用半乳糖，这个 Gal^-就是突变体。

一、突变的类型和它们的标志

突变可用几种方法来分类，一种方法是根据碱基改变的数目，仅有一对碱基的改变叫点突变，一个点突变可以是一个碱基对的替换、一个碱基的增加或减少。但点突变这个术语常常是指一个碱基对的替换。

第二种分类方法是根据所编码的氨基酸的顺序的改变，例如，如果编码的蛋白质中一个氨基酸被替换，这种突变叫误义突变（missense mutation），如氨基酸被替换后的产物——蛋白质在一种温度（典型的是 30℃）下有活力而在高温（通常是 42℃）下则无活力，这种突变叫温度敏感突变（temperature-sensitive 或 ts mutation）。根据惯例，把 Gal^-的温度敏感突变写为 Gal^-（Ts）。有些碱基被替换后就会使原来编码某一氨基酸的序列不再编码任何氨基酸，这种新的碱基序

列使蛋白质的合成在这一点终止，这种突变叫做链终止突变（chain termination mutation）或者叫无义突变（nonsense mutation）。无义突变有三种，每一种对应于一种三核苷酸，它们分别是琥珀（amber Am）、赭石（ochre O^c）和半透明(opal)。一个 Gal^- 的琥珀或赭石突变分别写成 Gal^-（Am）或 Gal^-（O^c）。

一个实验中如果分离了很多突变株，比如说分离了很多 Gal^- 突变株，那么可以把每株编上号，如 Gal4、Gal182（Am）、Gal123（Ts）等。

突变株的基因型常用某一特性的英文名称的前三个小写字母表示，表型也用前三个字母表示，但第一个字母要大写。例如色氨酸的基因表型用 Trp^+ 表示，合成色氨酸由几个酶完成，不同的酶在 Trp 后边再加一个字母如 TrpA、TrpE、GalK、GalT 等。对 TrpA 的突变株 103 则可写成 TrpA103。

二、突变株的筛选

检查突变株的方法有多种，而对于细菌突变株可以利用它们是否抗链霉素或能否利用某种糖作碳源来检测。

一个链霉素抗性菌（streptomycin-resistant Str-r）很容易在含有链霉素的培养基中被选出来。因为只有发生了突变的菌株才能在此培养基上生长。对于能利用某种糖作碳源的突变株可以用其在选择培养基平板上形成菌落的颜色来筛选。如不能利用乳糖作碳源的突变菌株（Lac^-）在 EMB 平板上形成白色菌落而野生型则形成紫红色菌落。

糖利用能力的突变型可以在缺乏这种糖而添加其他碳源的培养基上生长，而对于某种氨基酸的营养缺陷型如亮氨酸（Leu^-）突变株就不能利用上述在缺少氨基酸的培养基上生长的方法来筛选。因为这种突变株在这种培养基上并不能生长。对于这种突变株的筛选必须采用平板复印技术，选出的在基本培养基上不能生长而在完全培养基上能生长的菌落为氨基酸突变株。平板复印技术也可以用于 Ts 突变株的筛选。噬菌体突变株可根据噬菌斑的形态来鉴定。温度敏感突变株可根据在低温下能形成斑而在高温下不形成斑来鉴定。

在一个细菌的群体中，突变体的数目比未突变的数目要低得多，两者比例只有 10^{-7}，因此要从这么多的菌中分离出一株突变体是很难的。假设一个平板上长 200 个菌落的话，那么从 10^7 个细菌中分离一个突变体就需要倒 5×10^4 个平板（直径 9cm 平皿），这样分离一个突变体的工作量太大。那么如何才能提高分离效率呢？可以采取杀死野生型而不杀死突变型的培养基来富集突变株。一种适合于细菌突变株筛选的方法是青霉素筛选技术，这个方法主要的根据是青霉素对细菌细胞壁的合成的干扰，当在培养基中加青霉素时，野生型细菌因不能合成细胞壁而细胞内其他物质均能合成，结果细胞会破裂而死亡，而氨基酸缺陷型如 Leu^- 由于培养基中无亮氨酸使细菌不能合成蛋白质，因此 Leu^- 菌在青霉素培养

基上不会由于不合成细胞壁而胀破死亡。把含有 Leu^- 的菌悬液先在含青霉素的培养基上培养使野生型菌先死亡，然后再转入无青霉素而加亮氨酸的培养基中培养，这样便可大大富集 Leu^- 突变体。

三、诱　变

突变体的产生需要碱基顺序发生变化，这种变化可以由于 DNA 复制的错误而自发地产生，或者由以下五种途径促进产生：①插入的不能正确配对的碱基未被除去；②插入的一个异构化的碱基在接下来的复制过程中可以发生替换；③先插入的碱基发生化学变化成为具有不同配对特性的碱基；④在复制期间一个或多个碱基的缺失或插入。下面我们将分别讨论诱变剂及部分诱变机制。

诱变剂有以下几种类型，表 5-1 说明了各类诱变剂的作用。

表 5-1　诱变剂的类型

诱变剂	作用模式	例　子	结　果
碱基类似物	在复制期间替换一个标准碱基，并且引起在下一代的子代细胞中出现一个新的碱基	5-溴脱氧尿嘧啶 (5-bromouracil) 2-氨基嘌呤 (2-aminopurine)	A·T→G·C 和 G·C→A·T A·T→G·C
化学诱变剂	以化学方法改变一个碱基，从而引起在下一代的子代细胞中出现新的碱基配对	亚硝酸 (nitrous acid) 羟胺 (hydroxylamine) 甲基磺酸乙酯 [ethyl methane sulfonate (EMS)] NNG 紫外线 (ultraviolet light)	G·C→A·T 和 A·T→G·C G·C→A·T G·C→A·T G·C→C·G 和 G·C→T·A 和 EMS 同 所有单个碱基对的改变都是可能的
插入试剂	增加或减少一个或多个碱基对	吖啶类(acridine)	移框
增变基因	过多插入不正确碱基或者对插入错误碱基的修复缺失	—	任何单个碱基对的改变都是可能的
无诱变剂	5-甲基胞嘧啶的自发脱氨基作用 [spontaneous deaminatian of 5-methylcytosine (MeC)]	—	G·MeC→A·T

1. 碱基类似物诱变剂诱变

碱基类似物作为一种非标准核苷酸底物在 DNA 聚合酶作用下被组建到 DNA 分子中去，这种底物必须能与正在复制中的互补链上的碱基配对，聚合酶的

3′→ 5′校对功能可能将它们除去，然而这种类似物如果能够异构化或者有两种氢键模式，它将是诱变剂。

5-溴脱氧尿嘧啶（BU）是胸腺嘧啶（T）的结构类似物，它具有的溴和胸腺嘧啶上的甲基一样具有范德华力（图 5-7）。

(a) 胸腺嘧啶　　5- 溴脱氧尿嘧啶（酮式）

(b) 腺嘌呤　　胸腺嘧啶　　(c) 鸟嘌呤　　5- 溴脱氧尿嘧啶（醇式）

图 5-7　5-溴脱氧尿嘧啶的不同形式下的配对性能

在 DNA 复制过程中，BU 可以像 T 一样与 A 配对。BU 有两种形式，当处于酮式时类似于 T，与 A 配对；而当它处于烯醇式时（较少见）类似于 C，能与 G 配对。

如果 BU 处于烯醇式，DNA 第一次复制时 BU 以烯醇式掺入与 G 配对，因此最终使 C:G 对变成 A:T 对（图 5-8）。

当第一次复制时掺入的 BU 为酮式，则可使 A:T 对变成 G:C 对（图 5-8）。

无论 BU 是引起 A:T→ G:C 的变化还是引起 G:C→ A:T 的变化都要在 BU 掺入 DNA 分子后再经两次复制才会发生，因为只有 BU 在 DNA 复制过程中配对性能发生改变才能在子代中产生新的碱基对。

以上所说的碱基对的改变无论是 A:T→G:C 还是 G:C→A:T，都是由嘌呤-嘧啶到嘌呤-嘧啶的变化，这种由一种嘌呤-嘧啶变为另一种嘌呤-嘧啶的变化叫做碱基的转换（transition），如果是嘌呤-嘧啶转变为嘧啶-嘌呤的变化则叫做碱基的颠换（transverstion）。

碱基类似物除了 5-溴脱氧尿嘧啶以外还有 2-氨基嘌呤，2-氨基嘌呤是腺嘌呤的结构类似物，它能和胞嘧啶及胸腺嘧啶配对（图 5-9）。2-氨基嘌呤（AP）也能引起两个方向的转换，不过主要是引起 A:T→ G:C 的变化（图 5-10）。

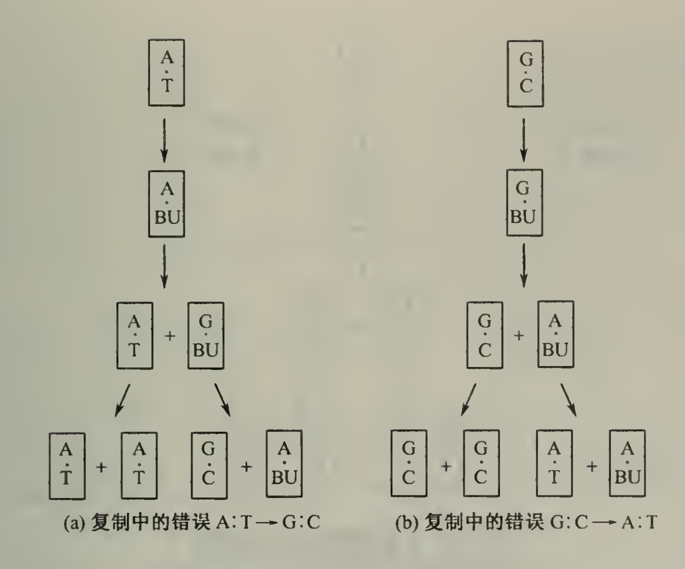


图 5-8　5-溴脱氧尿嘧啶引起 A:T→G:C 和 G:C→A:T 的诱变机制

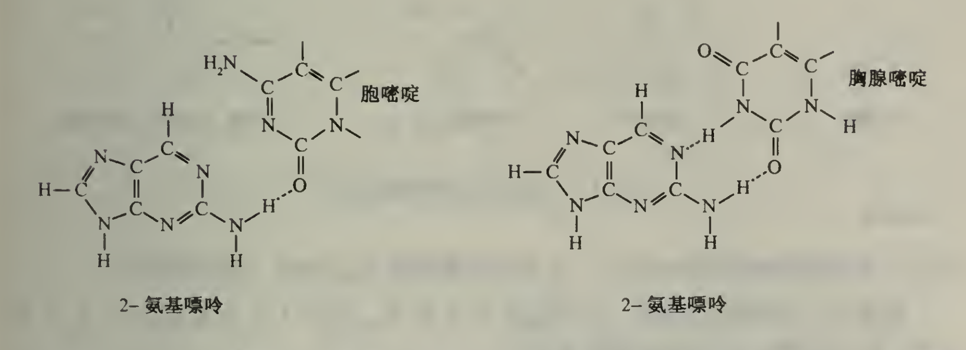


图 5-9　腺嘌呤的类似物 2-氨基嘌呤与胞嘧啶和胸腺嘧啶的配对模式

2. 化学诱变剂诱变

化学诱变剂能够改变 DNA 分子中的碱基形成氢键的特性。四种最有效的化学诱变剂是亚硝酸（nitrous acid）、羟胺（hydroxylamine HA）、甲基磺酸乙酯（ethylmethane sulfonate EMS）和 N-甲基-N 硝基-N-亚硝基胍（NNG）。它们的化学结构如图 5-11 所示。

首先亚硝酸通过氧化脱氨作用将氨基转变为酮基，这样，胞嘧啶（C）、腺嘌呤（A）和鸟嘌呤（G）经脱氨基就可以转变成尿嘧啶（U）、次黄嘌呤（H）和黄嘌呤（X），这些碱基可以形成碱基对 U:A→ H:C 和 X:C。因此 C 和 A 分别脱氨基后会引起 G:C→ A:T 和 A:T→ G:C 的转变。由于 G 和 X 都能和 C 配对，

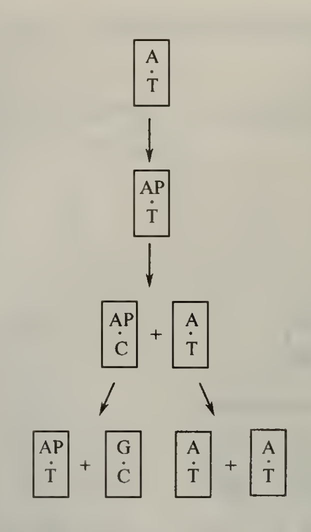


图 5-10　2-氨基嘌呤的诱变机制

亚硝酸　　羟胺　　甲基磺酸乙脂　　N- 甲基 -N 硝基 - 亚硝基胍

图 5-11　四种化学诱变剂的结构

所以 G→ X 的转变不引起突变。这些改变都是碱基的转换（图 5-12）。

羟胺专一地修饰胞嘧啶，使它仅能与 A 配对，所以 HA 引起 G:C→ A:T 的转变，这种转变的化学机制是复杂的。

EMS 和 EES 都是烷化剂，它们可以与 G 甚至有时也与 A 反应，将烷基加到嘌呤环的 N-7 位上并在这个位置上形成季胺，促进此环离子化（如图 5-13），离子化的烷基鸟嘌呤能与 T 配对而不与 C 配对。这样就引起 G:C→ A:T 的转换。另外，烷基嘌呤也有 N-糖苷键，很容易水解产生脱嘌呤位点，因此在 DNA 复制时可以插入任何一个碱基，原来的 G:C 对在复制两次后可能变成 C:G 对、A:T 对或者 T:A 对（如图 5-13）。

NNG 是一种非常有效的诱变剂，它起烷基化作用，对任何一个基因的诱变效率都可达 1%，用 NNG 诱变细菌常常会引起多基因突变。例如 NNG 诱导的 Gal^- 突变体常常也是 Pro^-，突变常分布在相邻的基因。对这种现象研究后发现 NNG 在复制叉处发挥最大效力，因此细菌在接触 NNG 时，复制叉处易发生突变。由于诱变处理都是在非营养的缓冲液里进行，DNA 复制很慢，所以突变的

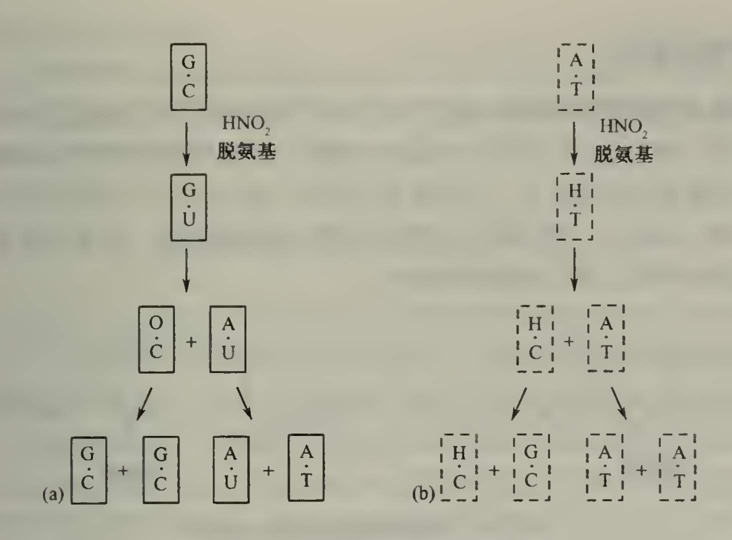


图 5-12　亚硝酸诱变的两种机制

图 5-13　烷化剂的诱变机制

基因常常是一个接一个。NNG 的优点是易找到突变体，但它的缺点是分离的一个突变株常常带有不希望的另—种突变。

3. 紫外线诱变

紫外线照射引起 TT 二聚体或其他嘧啶二聚体的产生。RecA$^-$ 细菌经紫外线诱变后不产生突变体，因为 RecA$^-$ 细菌经紫外线照射后不进行 SOS 反应，通常认为紫外线诱变主要是由于 SOS 修复过程引起的。

4. 插入类物质诱变

吖啶橙、前黄素和吖啶黄（图 5-14）等“吖啶类药物”都是三环平面分子。它们可以插入到DNA分子的两个碱基对之间，将两个本来相邻的正常碱基对隔开，隔开的距离正好相当于一个碱基对的距离（图 5-15)。这类物质的插入常引起碱基的插入或缺失（图 5-16)，当 DNA 复制之前插入，就造成碱基的插入；如果正在复制时插入将引起碱基的缺失。

NH_2 — 前黄素 — NH_2　　　　$(CH_3)_2N$ — 吖啶橙 — $N(CH_3)_2$

图 5-14　两类吖啶类衍生物的结构

A · T　　　　A · T
G · C　吖啶→　[ ]
T · A　　　　G · C
C · G　　　　T · A
　　　　　　　C · G

图 5-15　插入吖啶分子后两碱基对分开

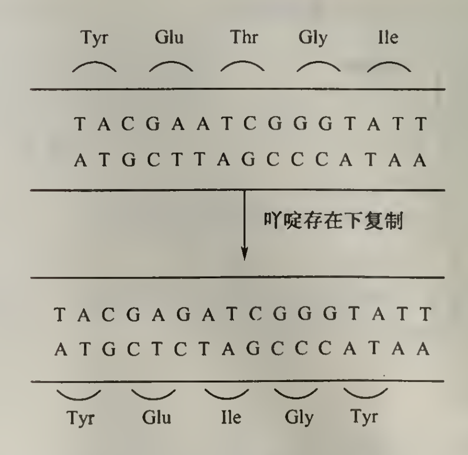


图 5-16　一个吖啶分子插入，复制时将增加一对碱基，因而引起密码改变

碱基的插入或缺失都将引起移框突变，也就是遗传密码的阅读框（reading

frame）向前或向后移一格。

除了上述物理和化学因素引起的突变以外，在 DNA 中插入可转移因子以及在大肠杆菌突变体状态中有一些基因引起突变的频率比遗传图上其他基因要高得多，这就是所谓的增变基因引起的突变。

四、Oligo 诱导的定点突变

用合成的 DNA 在 DNA 序列的特定位点诱导突变的方法始于 20 世纪 70 年代末期，这一方法受到 70 年代初用野生型 DNA 片段对 *ΦX*174 噬菌体的突变株进行标记拯救研究的启发，人们自然想到，既然可以用野生型的 DNA 片段使突变型 DNA 转变成野生型，当然就能用含突变的 DNA 片段使野生型 DNA 发生突变。到 80 年代中后期，这一方法随着 DNA 测序和 M13 系统、DNA 杂交等技术的发展得到较大的改进和提高，这些改进和提高主要有如下三点：①发展了一套用重组 M13 单链 DNA 为诱变模板的技术。②用标记的诱变引物为探针通过原位杂交法从平板中筛选变异的噬菌斑技术。③改进退火条件，提高野生模板与诱变寡聚核苷酸错配形成异源双链的效率。20 世纪 80 年代末期发展的方法实际上已经使人们可以随心所欲地在已知 DNA 序列中导入任意的突变，目前这些方法在世界上广泛采用。我们在初期也直接采用这些方法，但发现这些方法存在一些缺点，主要是：①产生变异的频率仍然不够高。用诱变反应混合物转染 TG1 感受态细菌，即使是技术熟练、经验很丰富的研究者获得的最好的结果，突变的空斑一般也只能在 40%～60% 范围，如果采用 JM107 菌株效率就更低，我们用 JM107 最高达到 10% 的变异率。此外，由于整套技术过程长、涉及的步骤多，对于条件稍差的实验室和经验不足的研究者来说，事实上很难获得所需的突变。②反应混合物直接稀释后转染感受态菌株，获得的空斑效率极低，从 0.5μg M13 模板出发获得的空斑数有时只有 2～3 个，情况好时也只有 10～20 个。③用诱变 Oligo 作探针通过原位杂交筛选变异的噬菌斑时，要定出一个非常严格的杂交温度，低 1～2℃ 时可以与野生型 DNA 杂交，高出 1～2℃ 时可以造成不杂交，尤其是只有一个核苷酸替换的情况下用杂交筛选就更加困难了。

针对上述问题，我们在历时两年多的对 Oligo 诱导的诱变研究过程中，对目前流行的 Kunkel 和 Zoller 的方法作了一些修改，效果相当明显，我们采用改进的方法对 HCMV DNA 聚合酶基因不同区域和 HIV U3R 启动子共进行了 10 种不同的诱变，其中 2～4 个核苷酸替换的突变频率可高达 80%～85%。诱导连续 4 个氨基酸缺失（缺失 12 个核苷酸）的突变频率也高达 40%～50%。由于突变频率很高，就可以直接挑选 3～5 个空斑，扩增后提取 DNA 进行序列测定来筛选，一般从 5 个空斑中可以获得 2～4 个突变。

下面将我们改进后的一套方法的步骤介绍如下：

1. 含 U 掺入的单链 M13 DNA 的制备

① 吸取出一个含目标 DNA 片段（即要进行诱变的 DNA 片段）的 M13 空斑，放入 0.5ml 的 2YT 中，涡旋振荡 10s，放 4℃冰箱过夜，次日离心 3min (10 000*g*)，吸出含噬菌体的上清。② 在 20ml 含 50μl/ml 尿苷（uridine）的 2YT 中接种一环 CJ236 菌，37℃培养过夜，然后在 1.5ml 含 50μl/m1 尿苷的 2YT 中接种 35μl CJ236 菌，并和 20μl 的上述 M13 上清混匀后静置 20min，37℃振荡培养 6h。③ 离心 3min 去除菌体，上清放 4℃保存作为毒种。这时如果用 CJ236 和 JM107 作空斑数比较，在 JM107 中存活的噬菌体数大约有 5%～10%。④ 在含 50μg/ml 尿苷的 2YT 中接种 CJ236 培养过夜，次日取 125μl 接种在 5ml 含 100μg/ml 尿苷的 2YT 中，再加入 50μl 含 UM13 的毒种，37℃振荡 6h。⑤ 离心去菌体，上清用于制备诱变模板单链 DNA。这样获得的上清中的 M13 在 JM107 中的存活率只有在 CJ236 中存活率的 1%～2%。

2. 诱变合成反应

① 退火：在一支 1.5ml 的塑料离心管中，加入 1μl 10×buffer（200mmol/L Tris pH 7.5 100mmol/L $MgCl_2$，50mmol/L NaCl，10mmol/L DTT)、0.5μg M13 模板 DNA、2～3ng 诱变 Oligo（模板与 Oligo 分子数比为 1∶10 以上，Oligo 经聚丙烯酰胺尿素凝胶电泳纯化以及 Kinase 反应在 5′加上磷酸基)，补水至 10μl。于 65℃保温 5min，转移到 37℃水浴 5min，取出放室温 10min，然后插入冰中 5min，离心 5min，加入 10×buffer（100mmol/L $MgCl_2$，100mmol/L DTT）1μl，dNTP 混合物（每种 2.5mmol/L）4ul（最终浓度每种 500μmol/L)，Klenow 片段 2.5μl，补水至总体积 20μl，混合后离心 5s，插入冰中 10min，放入 16℃水浴合成 8h。② 取出插入冰中继续维持 8h。

3. 酚-氯仿抽提

在反应混合物中加入 10μl TE，用酚-氯仿抽提一次，加入 1/10 的 3mol/L 乙酸钠（pH 5.2）和 2 倍体积冷乙醇，−70℃放置 30min，离心 10min，沉淀溶于 TE 中。

4. 转化

用 JM107 细胞转化，平皿空斑法同前。

5. 筛选和鉴定

挑选 3～5 个空斑分别扩增，提取 DNA 作序列测定。

上述方法的改进处在于：①含 U 掺入的 M13 诱变模板 DNA 制备。使野生

型 M13 模板 DNA 在 JM107 细胞中的存活率比原方法降低为原来的 1/3，这样就使得最后转染产生的空斑中野生型空斑的比例显著降低。②诱变合成反应中，采取分级退火、在冰中巩固异源双链结构和在冰中预先合成，后加 T4 DNA 连接酶等措施，可以提高诱变 Oligo 与模板 DNA 形成异源双链的效率，同时也减少 T4 DNA连接酶对 Klenow 片段活性的干扰从而提高 Klenow 片段合成 DNA 的能力。③把直接用反应混合物转染改为经酚-氯仿抽提后再进行转染，由于去除了 DNA 结合蛋白 Klenow 片段和 T4 DNA 连接酶，转染效率提高了 20 倍以上。我们曾把合成反应的混合物分成 2 份，一份按照原方法直接稀释转染 JM107，只获得 2 个空斑；另一份用酚-氯仿抽提后转染 JM107 获得 78 个空斑。

我们在进行 HCMV DNA 聚合酶的结构与功能分析中，用上述改进的方法作了 9 个不同位置、不同形式的突变都获得了很好的效果，发现并证实了酶的一些重要功能位点。诱导 DE80 突变，缺失 4 个连续氨基酸密码（12 个核苷酸），随机挑选 5 个空斑，通过 DNA 测序证实有 2 个突变，频率为 40%。DE96 诱导 9 个连续核苷酸缺失，突变频率为 50%（随机挑选 4 个空斑，测序后鉴定出 2 个突变），如果诱导 1～2 个核苷酸替换，突变频率可达 85%以上，诱导 HIV U3R 的 4 个核苷酸替换，随机挑选 10 个空斑，经 DNA 测序确定有 8 个突变。实践应用证实我们改进的方法与原方法相比可提高一倍以上的效率，使初学者也容易获得所需的基因突变。

第五节　基因突变的校正

所谓校正或抑制（suppression），就是使突变了的基因在一定条件下表达。这和恢复突变在本质上不同，它没有改变基因型，基因表达的条件是存在相应的抑制基因。

一、无义突变的校正

具有校正（或抑制）基因的细胞产生一种饰变了的 tRNA，这种 tRNA 可以识别无义密码子，并将无义密码子翻译成某种氨基酸，从而使多肽链的合成得以继续进行。

细胞中的抑制基因并非细胞原来固有的，而是另一种突变。它不是编码某种蛋白质的基因发生了突变，而是编码 tRNA 的基因发生了突变，使得原来不能识别无义密码子的反密码子环变成可以识别无义密码子的反密码子环。

例如，大肠杆菌中有一个菌株具有抑制基因 Su_3^+，它能使对应于一个无义琥珀突变的位置上出现一个酪氨酸，从而使蛋白质的合成能够通过无义密码子继续进行。酪氨酸的密码子是 UAC，酪氨酸 tRNA 的反密码子是 AUG，如果酪氨

酸 tRNA 基因发生突变使反密码子变成 AUC，它就能识别琥珀密码子 UAG，从而把一个酪氨酸放到 UAG 位置上，只要这一氨基酸的替代对于酶的活性影响不大，这一 tRNA 基因突变型就是琥珀突变型的一个抑制基因。

按照置换突变只改变一对碱基的原理，可以设想在终止密码子位置上出现的这些氨基酸的密码子应该和终止密码子只差一个碱基，以前发现的大肠杆菌的抑制基因确实有这一特性，在抑制基因作用下，在终止密码位置上出现的氨基酸是丝氨酸（UCG）、谷氨酰胺（CAA、CAG）、酪氨酸（UAU，UAC）、亮氨酸（UUG）、赖氨酸（AAG、AAA）和色氨酸（UGG）这些氨基酸的密码子确实只与终止密码子相差一个碱基。

二、错义突变的校正

抑制基因产生的饰变了的 tRNA 能识别另一种氨基酸的密码子而将野生型氨基酸插入多肽链，因而使合成的蛋白质恢复野生型的活性。

例如，色氨酸合成酶基因的 A_{36}突变导致一种无活性蛋白质的产生，野生型的蛋白质在这个位置上应该是甘氨酸（Gly），而 A_{36}在这个位置上却插入了一个精氨酸（Arg），但当 A_{36}突变株与 Su_{36}组成部分二倍体以后，就可以在这个位置上仍掺入 Gly，因而恢复蛋白质的活性。

三、移码突变的校正

移码突变的抑制机制还不是十分清楚，但在沙门氏菌中曾发现若干能抑制移码突变的基因，它们对无义突变和错义突变均无抑制作用。可以这样设想：①突变了的 tRNA 的反密码子环不是三联体而是四联、二联或五联体，因而可以把正常的氨基酸插入到相应位置去。例如甘氨酸密码子 GGG 突变后成为 GGGG，而突变的 tRNA 的反密码子环是 CCCC。②饰变后的 tRNA 结构扭曲，占据少于或多于三联密码子的空间。③突变后的 tRNA 在翻译同种碱基序列时发生位移。

移码突变的几个抑制基因（用 *Suf* 表示）。

Suf A	Pro CCC	*SufD*	Gly	GGG
Suf B	Pro CCC	*Suf E*	—	
SufC	—	*Suf F*	Gly	

第六章　DNA 重组

目前，我们可以把 DNA 重组归纳为四种类型：位点特异性重组、同源重组、转座重组和不正常重组。这四种重组的特点可归纳如表 6-1。

表 6-1　四种 DNA 遗传重组的特征

类型	是否需要同源序列	是否需要 RecA 蛋白	是否需要序列特异性的酶
位点特异性重组	+	−	+
同源重组	+	+	−
转座重组	−	−	+
不正常重组	−	−	?

同源重组有两个基本特点，首先是发生重组的两个 DNA 分子必须有相当长的一段序列（大约几百个碱基对）是相同或接近相同的，也就是说，发生互相作用的两段 DNA 序列必须是同源的；其次，这种重组过程需要宿主某一基因产物起作用，在大肠杆菌中的这种基因称为 *recA*，DNA 配对需要 *recA* 基因的产物。类似的基因也曾在其他细菌中发现。所以，同源重组也常被称为依赖于 *recA* 的重组。其余的三种重组类型都不需要 *recA*，也不需要或仅需要极有限的同源序列。例如在 λ 噬菌体的位点特异性重组中，需要整合酶的作用和 7 个碱基对的同源序列；在转座重组中，某一 DNA 片段在一个 DNA 分子内不同位置间移动或在二倍体细胞中的不同 DNA 分子之间移动（这种现象称为转座）也不需要同源序列和 *recA*；不正常重组是指非同源 DNA 序列之间的交换，之所以简单地把它称为“不正常重组”，是由于人们对这种重组过程知之甚少。DNA 重组过程似乎是复杂的，这些过程在原核和真核细胞中必然是不完全相同的，目前我们对这些过程的了解还相当少，本章中我们将以讨论同源重组为主。

第一节　同源重组的机制

一、断裂重接和异源双链

用高倍显微镜对玉米减数分裂重组的早期研究表明，在整个过程中染色体是先断裂然后再重新装配的。后来发现噬菌体杂交中裸露的 DNA 分子之间也可以发生重组。重组可以发生在一个基因中甚至可以不丢失任何碱基对，因而推断重

组的机制一定是十分复杂的。两个巨大的 DNA 分子如何能平行排列并且配对？断裂重接如何能够在两个 DNA 分子的同一个位置上准确地发生？这些都是研究同源重组的主要问题。

用两种类型的 λ 噬菌体做实验，一种是 A^+R^-，其 DNA 以重同位素标记；另一种是 A^-R^+，其 DNA 以轻同位素标记，这两种噬菌体均是 int^-和 red^-突变型，失去了噬菌体本身的全部重组功能，但由细菌决定的重组功能还存在，这样就能专门用于测验细菌重组系统及其重组机制。在这个实验中，噬菌体和细菌都带有阻止 DNA 合成起始的突变，因而 DNA 密度的变化只能由 DNA 的物理交换来引起，而不能由 DNA 复制引起。用这两种噬菌体感染细菌，子代噬菌体只含有亲代的 DNA。用氯化铯平衡梯度离心法测定所有噬菌体粒子 DNA 的密度，发现重组噬菌体（A^+R^+和 A^-R^-）的密度范围从全重到全轻，说明重组粒子中含有来源于两个亲代的物质，也就是说在形成的重组子中发生了物理交换。如果在 A 和 R 之间还有一个 C 标志，即 $A^+C^+R^-$和 $A^-C^+R^+$，那么要得到 $A^+C^+R^+$重组子，必须在两条 DNA 分子的右端即 C 和 R 之间发生一次断裂重接。另一个实验是用 int^+red^- λ 噬菌体（细菌是 $recA^-$），也是在复制缺陷系统中进行。重组仅发生在 λ 噬菌体的附着位点（att）之间。这种重组噬菌体只分为两个不同的密度，这是由于每个 DNA 分子在 λ 噬菌体附着位点断裂重接的结果。

图 6-1 C 基因近旁发生断裂重接产生杂合子

上述实验中观察到的另一重要现象是在 DNA 上 C 基因的近旁频繁地发生断裂重接，产生杂合子，即噬菌体粒子同时带有 C^-和 C^+两个标志，也就是说杂合子双链 DNA 分子中有一个一条单链上是 C^+而与之互补的另一条单链上是 C^-的异源双链区（图 6-1），带有这种异源双链区的重组噬菌体必须在 DNA 复制受到抑制时才能被发现，因为只要 DNA 进行复制，这种异源双链就会分开，杂合子将不存在。

噬菌体-噬菌体和其他类型的重组实验确凿地表明断裂重接是同源重组的所有模型中一个必需的步骤，并且异源双链区结构常常是中间体。图 6-2 表示了在酶作用下的典型的重组模型。这个模型的机制虽然还不十分清楚，但概述了 DNA 交换的一系列特征。

二、支链迁移

在第二章中已经讨论过 DNA 分子中存在泡泡（breathing），支链迁移是在 breathing 基础上，双螺旋区中的一条碱基配对的链被另一条也是配对的链代替的现象。如图 6-3 所示，w 链是一条连续的分子，两条单链 I 和 II 都和 w 链互补，它们和 w 链结合的长度超过了 w 链。如图 6-3 中单链 I 和 II 中的粗线部分

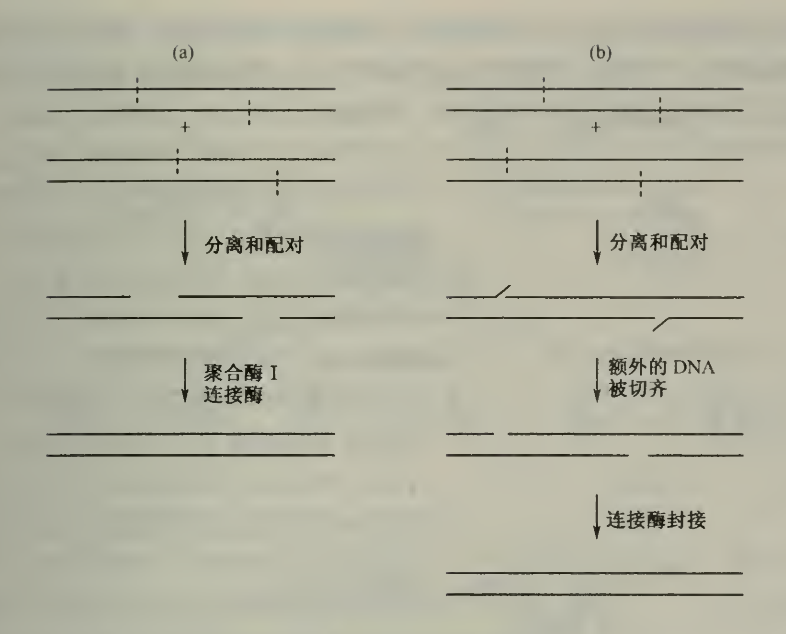


图 6-2 断裂重接的简单模型

都与 w 链中的虚线部分互补。当 breathing 发生在右端时 I 链迅速与 w 链分开而 II 链则与 w 链结合［图 6-3（c)］，breathing 可再次发生，根据其发生的位置以及两条链与 w 链的结合可能出现图 6-3 中（a)、(b)、(c）等不同情况。在大部分重组模型中支链迁移起重要作用。

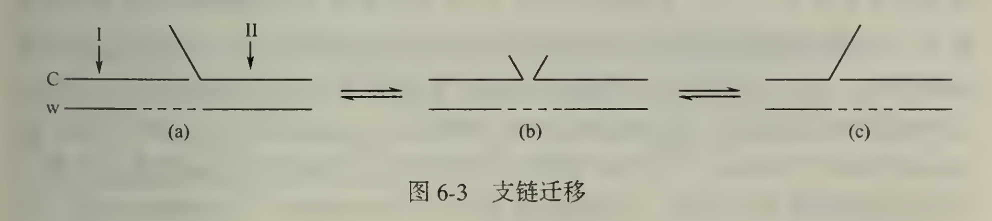


图 6-3 支链迁移

三、碱基对的错配及消除

异源双链区含有对应于野生型位点的突变而产生错配的碱基对。这样的错配是不能持久的，它们将被复制或错配修复所消除。

错配修复最初是在 λ 噬菌体的实验中发现的（图 6-4)。从两种 λ 噬菌体突变体中提取 DNA，变性后分离纯化互补单链，将含有一个突变等位基因的单链和含有另一突变等位基因的互补单链复性，用此复性 DNA 去转染重组缺陷型细菌，培养后再检查被感染细菌中的野生型噬菌体。转染后复制应产生数目相等的

两种单突变噬菌体，但是，如果在复制之前发生了错配修复，那么将会产生一些野生型噬菌体。产生野生型噬菌体的细菌有三类：含有野生型和 a 突变，含有野生型和 b 突变和只含有野生型噬菌体（图 6-4）。

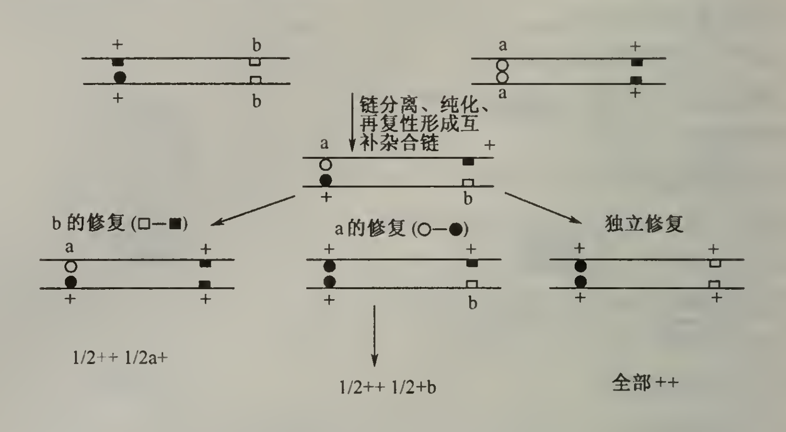


图 6-4　错配修复实验

错配修复是通过切除异源双链区上的一个单链片段来完成的。我们可以根据图 6-4 中的各种实验来估计在错配修复过程中切除了的碱基对数目。实验中引入了第三个标记并且与前两个标记之一紧密连锁，然后检测被感染的单个细胞中的子代噬菌体，看第三个标记的修复是否常常伴随着与其紧密连锁的标记的修复，即看具有 $a^+b^-c^-/a^-b^+c^+$ 基因型的异源双链子代 a^+b^+（b 和 c 是紧密连锁的）是否也常常是 $a^+b^+c^+$（异源双链中的 b^- 改正为 b^+），发现共改正（b^-c^- 改为 b^+c^+）的频率依赖于两个标志的间隔和它们的相对位置。对一系列标记的检查结果表明，当在一个特定的错配位点起始错配修复时，修复将沿着此单链以 5′→3′方向继续进行，延伸大约 3000 个碱基对。现已发现在大肠杆菌中有 Mut H、Mut S 和 Mut L 三种蛋白质参与错配修复，具体修复过程见第五章第一节中有关错配修复的内容。在对错配修复的研究中观察到一个有意思的特征，即在对不同标记的实验中观察到了改正的方向选择，对于某些标记来说，错配修复将野生型改为突变型，而对另一些标记来说则是将突变型改为野生型。这个现象还未得到解释，也许错配碱基周围的碱基序列决定了这个链中的切割位置。

四、DNA 分子的配对

DNA 分子的配对对于所有同源重组模型都是必需的。1965 年分离到了一株 *E.coli* F^- 突变株，它不能作为 Hfr 的受体，突变基因定为 *recA*，1976 年分离出了 RecA 蛋白。

RecA 蛋白有两个主要的生物学活性：①与单链 DNA 结合；②具有蛋白水解酶活性。与单链 DNA 结合是它与重组有关的一个特性，而另一活性是起调节作用。当 RecA 蛋白起 DNA 结合蛋白的作用时可以介导 DNA 分子的非专一性配对和同源依赖性链入侵。用纯化的 RecA 蛋白和 DNA 分子在室温下作用，二者由于互补碱基 AT、GC 配对而保持在一起，这种结构是很稳定的。对图 6-5 中所示的三种互相作用的研究以及对没有互相作用的 DNA 分子配对的研究表明，稳定的配对依赖于两个条件：①DNA 分子必须是单链或具有单链区；②两个分子都必须具有自由末端。第一个条件常由缺少自由末端的超螺旋 DNA 提供，第二个条件排除了单链环入侵双链环的可能性。RecA 介导的这种相互作用分三步完成：RecA 蛋白与联会前的单链 DNA 结合、联会以及联会后的链交换（图 6-5）。这些步骤是通过最初对噬菌体感染的细胞中双链环状 DNA 分子与噬菌体的单股正链之间的配对研究得到阐明的。

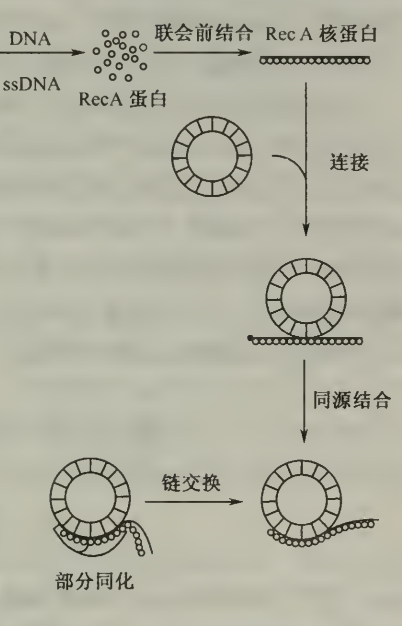


图 6-5　RecA 介导的配对

1. RecA 蛋白在单链 DNA 上结合

如果将 RecA 蛋白和单链 DNA 混合，RecA 蛋白将覆盖此单链 DNA 形成核蛋白细丝，此结构呈现带状，表明它有很大的规律性。每条带含有 6～8 个 RecA 蛋白分子，对应于 20 个核苷酸。这个被 RecA 蛋白覆盖的 DNA 分子的长度是它在双链状态时的长度的 1.5 倍。

2. 联会

在 ATP 作用下，核蛋白和双链环状 DNA 分子形成复合物，最初的相互作用不是在同源区之间进行的，非序列依赖性作用之后，两个 DNA 分子相对移动直到同源序列接触到一起，这叫同源结合。当同源序列排列在一起时，这两条链也不缠绕在一起。

3. 联会后链交换

同源排列但互不缠绕的两条链只是很弱地结合在一起，这种结构是很不稳定的。在体外，*E. coli* 拓扑异构酶 I 能使之形成正确缠绕的双螺旋，但不知道在体

内是否也是这样。一旦在单链的一端形成同源排列区，RecA 蛋白就积极地启动双链分子中一条链的替换和新链的同化，这个反应的机制还不清楚，但 RecA 起解旋酶（helicase）的作用这一点已经很清楚了，它在形成异源双链的进程中能解开双链 DNA 的螺旋。同化过程是一种极性过程，移动只以 5′→ 3′的方向进行。

我们在图 6-5 中看到的每一种相互作用的过程（单链 DNA 与双链 DNA 的三种作用过程）都可以用上述 RecA 介导的多步过程来解释。但这似乎与两个双链 DNA 分子之间的同源配对没有明显的关系，双链 DNA 分子的这两种配对发生在减数分裂、噬菌体与噬菌体重组、转导和接合过程中。所以，由 RecA 介导的同源重组中同时需要单链 DNA 和一个自由的末端。我们会看到，几乎所有的重组模式都包括一个早期步骤，在这一步骤中，一条 DNA 链被切出一个缺口，并且以不同的方式从缺口处解开螺旋。

第二节　细菌转化中的重组

一、细菌中的转化

前文所述的 RecA 介导的单链入侵结合、支链迁移和错配修复也同样构成了细菌转化中的重组机制。外源 DNA 和染色体 DNA 重组在肺炎双球菌的转化中了解得最早也最为清楚。关于肺炎双球菌的转化过程及其在分子生物学中的重要意义前面已经论述过，现将细菌转化中的重组模式以图解表示（图 6-6）。从图

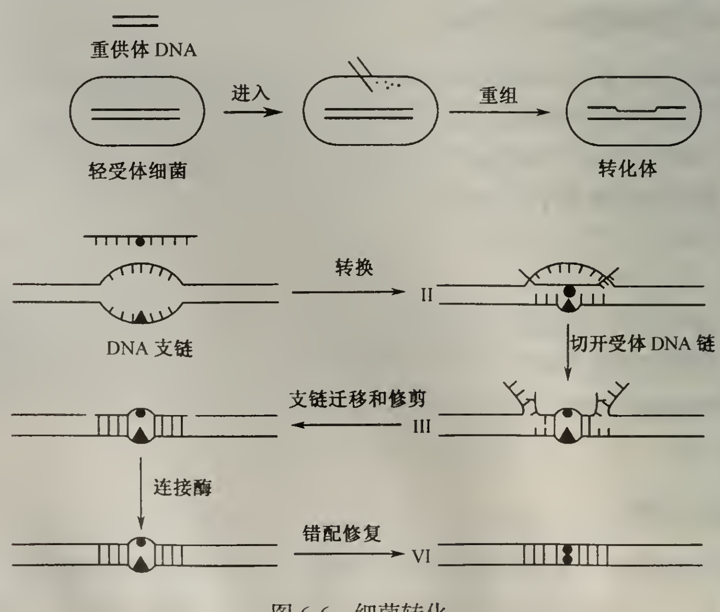


图 6-6　细菌转化

中可以看出外源 DNA 被转入细胞中以后，其中一条链被降解，只有一条链与受体 DNA 相互作用，后者为什么能够抵抗核酸酶的降解目前还不清楚，很可能是这条单链被单链结合蛋白结合而受到保护。单链入侵的过程与 RecA 蛋白介导的配对一样，首先是切断被替换的单链，然后经支链迁移、修剪配对除去自由末端，最后由连接酶连接切口，产生一个含有错配区的双链。供体的标记能否在受体中出现要根据错配修复能否发生而定，如果错配修复能发生，那么要看究竟是供体的还是受体的碱基被除去。对于一些高效能的标记，错配既能被改正为与供体基因型相匹配，而且也很少被修复。在不被修复的情况下，随着细胞的分裂会分别产生一个供体基因型和一个受体基因型的细胞。由于选择培养基只允许重组体生长，所以形成的菌落只由重组体组成。对于低效能标记来说，错配修复常除去来自于供体的错配碱基，而细胞只保留受体基因型。自从在肺炎双球菌中发现转化以来，在其他种属细菌中也发现了转化，如流感嗜血杆菌和枯草杆菌等，但还有许多其他种属的细菌不出现转化。在能进行转化的细菌中也只有少数菌株可以作为有活性的转化受体，而这些菌株也只有在生长状态的一定阶段才变为可转化状态，即所谓的感受态。

一定的外界因素也可以使通常不具有感受态的受体表现感受态。例如使用 Ca^{2+} 和改变温度可以使 *E. coli* 细胞更容易接受外源 DNA。又如聚乙二醇 6000 可以促进原生质体摄入 DNA，这种技术可用于重组 DNA 的许多类型的转化，但必须在原生质体重新形成细胞壁并在培养基中形成菌落时才能用于转化实验。用于转化的外源 DNA 分子的分子质量要比较大，分子质量小于 10^6Da 时转化效率很低，一般要达到 10^7Da 左右，DNA 分子的转化效率才比较高，此外双链 DNA 分子的转化效率也比较高。来自质粒的双链环状超螺旋 DNA 分子也可以使感受态细胞、Ca^{2+} 处理细胞和原生质体转化，但必须具备自己的复制原点。

关于感受态的本质有两种假设，即局部原生质化假设和酶受体假设。这两种假设都有实验证据，特别是酶受体假设，由于已经分离到了感受态酶而得到有力证明。感受态的出现与细菌的代谢、生长阶段和遗传背景有关。

二、酵母中的转化

酵母中转化的机制与肺炎双球菌中的十分不同，曾用一种带有一小段酵母 DNA 的 *E. coli* 质粒作为供体，酵母细胞作为受体来研究酵母的转化。酵母 DNA 缺少复制原点，而细菌的复制原点在酵母中是失活的，因此在 *E. coli* 质粒中的那段酵母 DNA 要在酵母细胞中表现必须发生重组，酵母转化的下述特点与其他转化系统是明显不同的。①整个质粒被掺入到细胞中，交换只发生在酵母 DNA 同源序列之间，交换的位置可以分布在同源区域中的任何地方。②酵母中的转化效率远低于肺炎球菌中的转化效率，但是如果用限制性内切核酸酶把质粒

中的酵母 DNA 序列切断，转化效率就可以提高几千倍；相反，双链断开的 DNA 在肺炎球菌中以及在所有已知的细菌转化系统中转化效率都会降低很多倍。③双链断开的 DNA 除了转化效率提高外，还产生了一种交换位点的倾向性。例如，如果质粒上含有两个酵母基因 A 和 B，当在质粒的 A 基因切断双链 DNA 时质粒 DNA 就被插入到酵母染色体的基因 A 中；如果在基因 B 中切开，质粒就插入酵母染色体的 B 基因中。在上述两种情况中都会使转化的酵母细胞具有两个基因 A 和两个基因 B，这种结果说明在重组过程中需要双链 DNA 断裂被修复和断裂附近发生交换。

下述实验使我们看到一些在酵母中发生重组的过程（图 6-7），在这一实验中，供体（质粒）DNA 带有 A^-B^+ 等位基因，酵母细胞中带有 A^+B^- 等位基因，在质粒 DNA 的 A^- 基因中切开，去除大部分 A^- 基因，使 A^- 产生大范围的缺失，这样的 DNA 和双链断裂的 DNA 一样可以使转化效率提高几千倍。值得注意的是在转化的细胞中不但含有两个 B 基因，而且含有两个 A 基因，两个 A 基因都是受体的 A^+，缺失的 A^- 等位基因消失了。这种基因倍增表明由两种酶切割产生的缺口被修复了，这种修复过程是一个以受体的 A 等位基因为模板的复制过程。实验中另一受体等位基因为负，质粒的等位基因为正，在转化的细胞中总是发现两个拷贝的受体等位基因，这种称为“双链断裂修复模式”被提出来解释酵母中的重组过程。这一模式的要点如下：①通过两次独立的复制过程修复双链断裂的 DNA，每次各复制一条单链。②在双链断裂的 DNA 修复的地方的两侧各形成异源二聚物区域。③形成两个 Holliday 交叉链的中间体。④通过解决两个 Holliday 接头而完成重组过程。

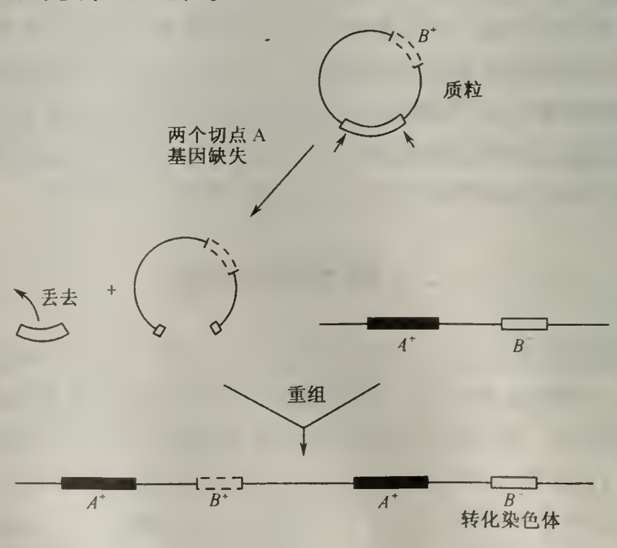


图 6-7　酵母中的转化

第三节　同源双链 DNA 分子之间的交换

目前已知道了单链 DNA 与双链 DNA 之间转化的交换机制。但双链 DNA 分子之间的交换更为普遍，例如整合酶启动的双链 DNA 中的专一位点的交换，这种专一位点交换的两个 DNA 分子大部分是不同源的（如 λ 噬菌的整合）。双链 DNA 分子之间的交换以同源交换类型为主，如细菌的接合转移、转导和减数分裂交换，但机制还不完全清楚。

一、噬菌体的整合

它是通过 λ DNA 上少数碱基对与宿主染色体 DNA 上的同源性，在噬菌体编码的整合酶的作用下发生的 DNA 双链之间的交换重组。整合模型如图 6-8，图中 PP′是 λ 噬菌体 DNA 的 att（attachment）位点，BB′是细菌染色体上的 att 位点。λ 噬菌体 att 位点的碱基顺序是：

GCTTTTTTATACTAA
CGAAAAAATATGATT

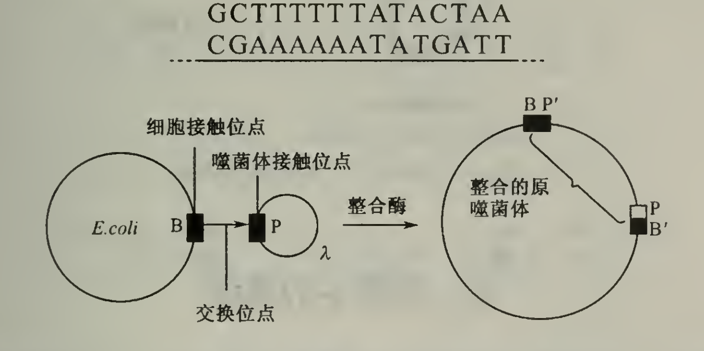


图 6-8　细菌噬菌体 λ DNA 的整合

可以看出这个顺序中 AT 碱基对非常丰富。关于 λ DNA 的整合机制还不清楚，但交换发生在 BB′和 PP′之间，交换后产生两个重组的位点 BP′和 PB′(图 6-8)。图中只表明了重组的方式，但并没有说明重组的机制。重组发生在核心部分（同源的 15bp），核心的两旁虽然没有显示顺序的保守性，但体外实验证明它们也可影响重组。用核酸酶逐渐缩短 att PP′和 BB′长度后再进行重组实验，结果证明 P 的必要长度是 160bp（0～160bp），P′的必要长度是 80bp（0～80bp），B 和 B′的必要长度都是 11bp。这些结果说明虽然重组发生在核心序列中，但是重组蛋白不只作用于核心序列。实验还证明一条链上的断裂重接位点是 +4，另一条链上是 −2。整合酶可以与 attP 结合并且具有拓扑异构酶 I 的性质，除整合酶

外，整合还需要宿主编码的宿主整合因子。根据整合酶的拓扑异构酶 I 的性质，可以设想两个 att 位点发生联会以后，在整合酶作用下两个 DNA 分子的每一单链断裂，在一瞬间旋转以后仍然在整合酶作用下连接成为半交叉（Holliday 结构，见重组模型），另两条单链以同样过程断裂重接，完成全部重组过程。

二、细菌接合转移中的重组

虽然许多实验证明供体菌 Hfr 与受体菌 F^- 进行接合转移过程中最初转移到受体菌中的染色体 DNA 是单股，但在与受体菌染色体发生同源重组之前已进行了复制，因此供受体 DNA 发生的重组也是双链 DNA 之间的交换重组（图 6-9）。

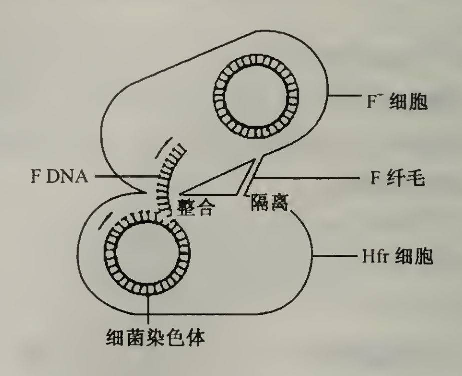


图 6-9　Hfr 和 F^- 之间的接合转移

三、转导中的重组

无论是专一性转导还是普遍性转导，都是双股 DNA 分子与染色体之间的重组，即双链 DNA 分子之间的重组。转导中的重组与转化中的重组不一样，在转导重组中染色体 DNA 双链必须打开，由于对转导中的重组机制了解得很少，不便进一步讨论。下边将对不同的转导类型特征及转导在遗传分析中的应用略作介绍。

转导分两种类型，一种是转导噬菌体，只能转导细菌染色体上某一特定基因片段，这种转导称为专一性转导（或称限制性转导），专一性转导最早由 Lederberg 发现，他用紫外光照射野生型大肠杆菌 *E.coli*（λ），将得到的裂解液感染 *E.coli* K12 的各种非溶源的突变株，发现只有 gal^- 突变株有 10^{-6} 的几率被转导成了 gal^+，其他突变体均不被转导成野生型，说明 λ 噬菌体只能转导供体菌染色体上特定基因，所以叫专一性转导。

为什么 λ 噬菌体只能进行专一性转导呢？这与 λ 噬菌体的溶源状态和整合割

离特点有关，λ 噬菌体整合位点附近刚好是 *gal* 基因，在某种条件下，整合的 λ 噬菌体可以从染色体上割离下来，割离可以是正常的也可以是错误的，正常割离后产生完整的 λ 基因组（图 6-10），错误割离的几率是 10^{-6}，错误割离后产生带有一段染色体基因（如 *gal*）同时缺失了 λ 本身一部分基因的 λ 噬菌体的基因组（图 6-11），这种由于错误割离产生的 λ 噬菌体就是转导型噬菌体，也称 λ dg。用这种方法得到的 λ 裂解液转导频率很低（10^{-6}），称低频转导（LFT）。如果用 λ dg和 λ 同时感染 *E.coli* K12，可得到双重溶源菌 λ dg/λ，诱导这种双重溶源菌所得到的裂解液中有 50%是转导型噬菌体，称为高频转导（HFT），双重溶源菌的形成如图 6-12 所示。

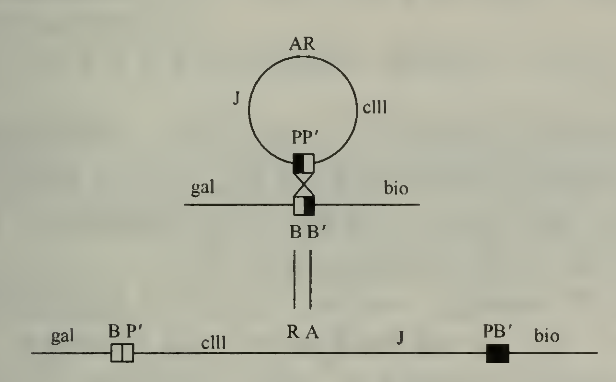


图 6-10　λ 噬菌体的整合与割离

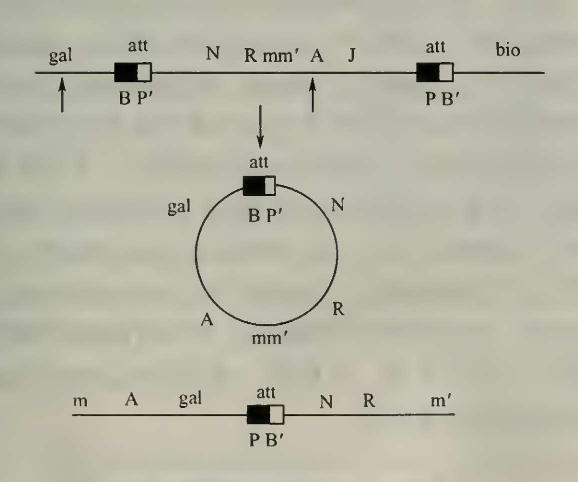


图 6-11　错误割离产生转导型噬菌体

普遍性转导是 Lederberg 和他的学生 Zinder 1951 年研究鼠伤寒沙门氏菌时发现的，他们的目的是研究鼠伤寒沙门氏菌 λ 噬菌体的整合与割离中是否存在接合

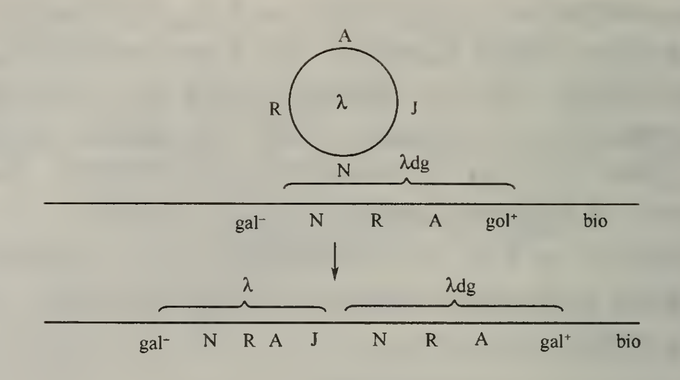


图 6-12　双重溶源菌的产生

转移现象，结果发现了这种菌中存在一种滤过性因子，能转移细菌的任何基因，后来证明这种滤过性因子就是温和性噬菌体 P22。由于 P22 能转导任一基因，所以称普遍性转导。除了 P22 外，P1 噬菌体也能进行普遍性转导，并且 P1 可以感染 *E.coli*，所以用 P1 来研究普遍性转导更为方便。

P1 和 P22 为什么能转导任一基因呢？这是因为 P1 是在细胞质中溶源（就像质粒和细菌共存一样）。P1 转导型噬菌体的形成是由于错误包装。当诱导 P1 溶源菌时，P1 噬菌体在宿主细胞中复制，将宿主染色体打成片段，包装时将与 P1 基因组大小一致的染色体片段包入 P1 头部，这样就形成了转导型的噬菌体，为了证明这一点，将 thy^- 菌分别在胸腺嘧啶和 5-溴脱氧尿嘧啶的培养基中培养。共三份：A 中加胸腺嘧啶，B 和 C 中加 5-溴脱氧尿嘧啶，培养后用 P1 去感染，其中 B 份中 P1 感染时转入胸腺嘧啶中培养，将裂解液在 CsCl 中密度梯度离心，达到平衡时分部收集噬菌体并且分析各组分中有侵染力的和有转导作用的 P1 颗粒。结果发现 B 份中有侵染力的密度与 A 份中的相同，而有转导能力的 P1 的密度与 C 份中的相同，这就充分证明了转导型 P1 包装的是宿主的 DNA。

P1 噬菌体的头部能够包装 *E.coli* 染色体上 2min 的长度，所以染色体上两个相距很近的基因可以同时转导，这叫共转导（cotransduction）。两个基因相距越近共转导频率越高，越远共转导频率越低，所以可以通过测共转导频率测定两个相距很近的基因之间的距离。根据这个原理 Yanofsky 和 Lennox 等绘制了 *E.coli* trp 区段的精细结构遗传图。

四、减数分裂重组

在子囊菌中，每一次减数分裂的产物包含在一个子囊中，而且以一定的顺序排列，因而用子囊菌作材料来研究减数分裂中的重组机制比较方便。

在 20 世纪 30 年代中期就出现了关于染色体交换的两种假设：断裂重接和复制选择模型。

按照断裂重接模型，两个染色单体在同一位置上断裂，然后彼此和另一染色单体重新连接起来。按照复制选择模型，两个染色单体作为复制的模板进行复制，在复制过程中相互交换模板，从而造成交换。

两个模型的重要区别是，断裂重接模型认为重组可以在 DNA 不进行复制时进行，而复制选择模型则认为重组必须在 DNA 复制时进行。

现在已知道染色体复制在间期已经完成，而交叉出现在减数分裂的早期，这一事实支持断裂重接模型。另外，用标记噬菌体进行实验，发现重组与 DNA 复制无关，也支持断裂重接模型。

在粗糙链孢霉等子囊菌中，杂交子代的子囊有六种类型：AAAAaaaa，aaaaAAAA，AAaaAAaa，aaAAaaAA，AAaaaaAA 和 aaAAAAaa，而且每一对孢子的基因型总是相同的。后来在面包酵母和粪生粪壳菌等菌中发现一些不正常的分离现象。在粪生粪壳菌的野生型黑色子囊孢子和突变型灰色子囊孢子的杂交（$g^+ \times g^-$）中，曾观察了 200 000 个子囊，其中绝大部分属于正常的 4∶4 类型，不正常的分离类型包括：5∶3120 个，占 0.06%；6∶2100 个，占 0.05%；不规则的 4∶416 个，占 0.008%。

图 6-13A 表示一个不正常的 4∶4 子囊，其中 g^+ 和 g^- 的分离是不正常的，A 和 a 的分离是正常的。图 6-13B 表示一个 5∶3 不正常分离子囊，其中从上往下数第三对孢子对于接合型来说是正常的，可是对于 g 来说则是不正常的，说明在这里分离发生在减数分裂后的有丝分裂过程中，所以称为减数分裂后分离。一个等位基因的 6∶2 和 5∶3 分离也叫基因转换。对几种真菌和果蝇的不同性状的观察使人想到基因转换过程与减数分裂后分离及交换是相互联系的，这种关联是同一步骤所产生的结果，即单链多核苷酸之间的配对和交换，并且形成异源双链区。

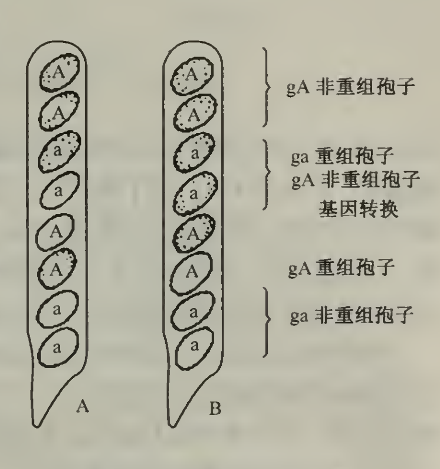


图 6-13　不正常分离子囊

异源双链区的形成可用来解释减数分裂后分离，也可以用来解释转换，这两种现象的区别在于异源双链区形成的步骤。

异源双链区是由于交换造成两个仅一个 bp 不同的标志（如 b 和 B）而形成的，这样形成一个错配的碱基对（图 6-14）。①I 中没有错配修复，这样每个错配的链复制后产生一个带野生型等位基因的 DNA 分子和一个带突变型等位基因的

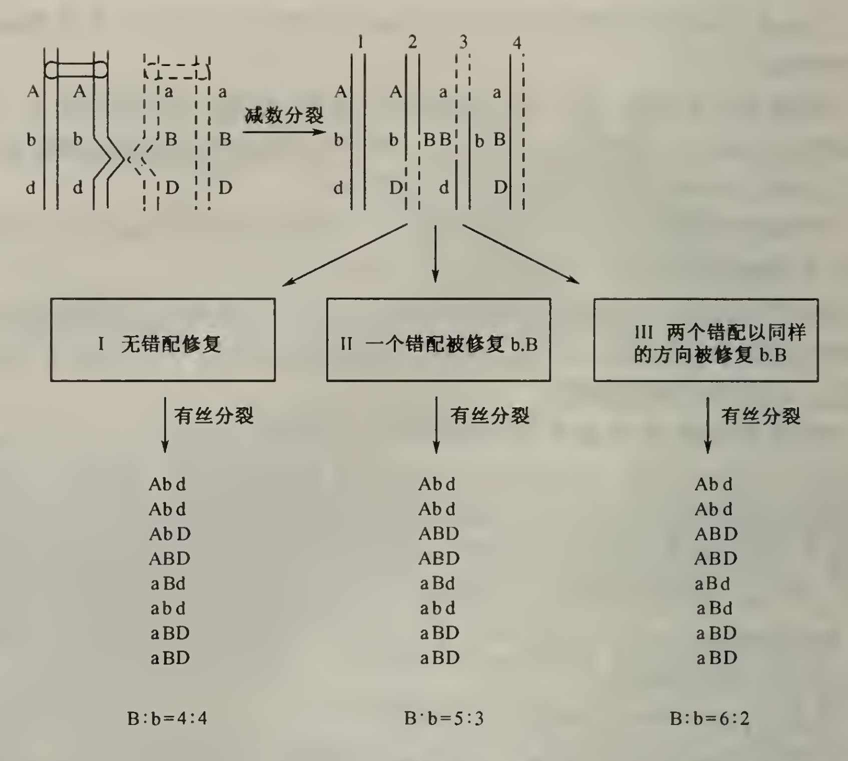


图 6-14 在三种不正常子囊形成中错配修复的作用

DNA 分子。因为两个分子带有错配碱基对，所以减数分裂产生两对含不同成员的孢子，B:b 是 4:4，但子囊是不正常的 4:4 类型。②II 中错配修复将分子 2 异源双链区中的 b 改正为 B（b 被切除，B 作模板），结果产生含不同成员的孢子对，B:b 是 5:3。③在 III 中两个错配碱基对被以同样的方向修复，即 b 改成 B，所以所有孢子对均含相同成员，B:b 是 6:2。

图中的不正常 4:4 与两旁标志重组的转换子囊相关联，另外图中有一半非正常子囊中的侧旁标志是非重组的，对于这些问题目前的解释如下：在 DNA 分子配对的某一步形成异源双链区，配对分子的分离需要在每条多核苷酸链中产生两个切口，这样可能发生两种排列方式，一种是产生侧旁标志重组子囊，另一种是产生非重组子囊。

第四节 同源重组模型

1964 年 Robin Holliday 的创新模型中首次提出形成的异源双链区随后被错配修复。Holliday 模型的一个重要推论是预言了一个结构，现已证明这个结构是存在的。现在 Holliday 模型已被修改多次，本节叙述的第二个模型是重要的修改模

型之一，虽然此模型的细节可能会因为所获得的更多的资料而被修正，但它总能很好地描述发生在真菌中的交换过程。

一、Holliday 模型

Holliday 模型和以后的修改模型都是为了说明正常子囊、基因转换和减数分裂后分离的原因和说明不正常子囊与交换的关系。Holliday 模型如图 6-15。模型中的步骤如下：①联会中的两条染色体（双链 DNA 分子）中的两条单链在同源位置被一种目前还不清楚的机制打断。②在每个断裂的地方双螺旋稍微解开，释放出单链。③释放的单链通过与另一分子中没有断裂的链互补配对而在两个 DNA 分子间进行交换，形成一个由于单链交叉而使两个 DNA 分子连结在一起的结构。④对称地产生切口，两个 DNA 分子分离。切口可以在两个不同的位置产生［图 6-15（b）中所示］，如果从箭头 1 和 2 处切开，产生（c）中的两个非重组分子，而如果从箭头 3 和 4 处切开，产生（d）中的两个重组分子。这种使两个参与重组的 DNA 分子分离的切割过程叫解离。

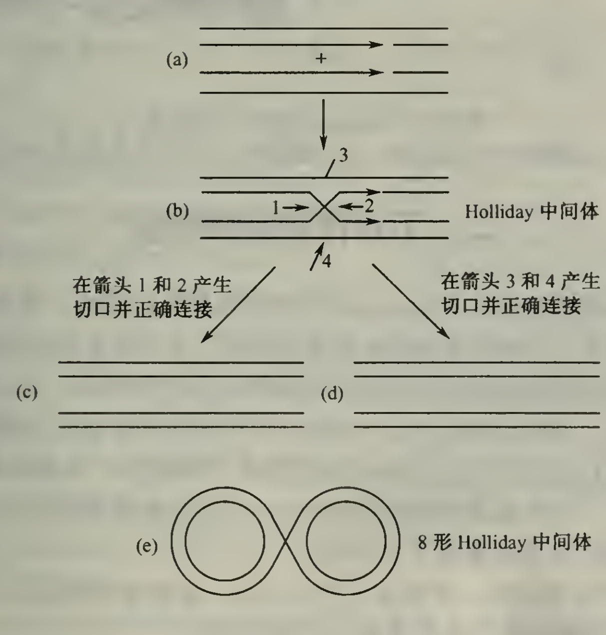


图 6-15　Holliday 模型

Holliday 模型为不正常分离和交换之间的联系提供了分子基础。基因转换和减数分裂后分离是由于含有异源双链区的 DNA 对称结构中的一条或两条染色体上发生错配修复或无错配修复而产生的。此模型已解释了在侧旁标志不重组的情

况下发生 5:3 和 6:2 的分离。如图 6-15（b）中，一个 Holliday 交叉通过切口配对可产生两侧标志重组或不重组的子囊，如果以东西方向切割（箭头 1、2 处），则产生侧旁标志不重组的子囊，如果切割以南北方向（箭头 3、4 处）进行，则侧旁标志重组。如果这两种切割大致相等，将可观察到有一半的 5:3 和 6:2 子囊的侧旁标志是重组的。

这种交叉区已在几种生物 DNA 中得到证实，这种结构叫 Holliday 交叉，也称为 Holliday 结构，当交叉以一个“开－盒”（open-box）构象存在时非常容易辨认，当 Holliday 结构的一个组成部分旋转 180°时，将产生“开－盒”式分子（图 6-16）。



图 6-16　Holliday 结构的不同构象

(a) 按箭头方向旋转后产生开盒结构。(b) λ 噬菌体 DNA 分子之间形成交叉的电镜照片。

二、不对称链转移模型

不对称链转移模型保留了 Holliday 模型的同源单链连接、存在异源双链区、产生 Holliday 交叉、切割分离和错配修复等特征，不同的是最初的配对是不对称的，主要是根据对酵母等研究的数据。在最初的相互作用中，只为受体 DNA 分子提供一个单链，异源双链区只在入侵的分子中形成。随着入侵链复制的扩大和支链迁移，异源双链区扩大并以新的形式重排（异构化），在两个分子中形成异源双链区。第二个区别是某些转换子囊可以不通过错配修复而形成。不对称链转移模型如图 6-17，具体步骤如下：

1) 两个分子中其中的一个单链上产生切口。该模型中假设只有一条链被打断，而在 Holliday 模型中，两个分子中都产生切口。

2) 替换（displacement）。在一种 DNA 聚合酶的作用下，切口 3′-OH 自由末端加上一些核苷酸替换了原来的多核苷酸，从而产生一单链分支，此分支用于下一步的互补配对。

3) 侵入（invasion）。可能在 RecA⁻ 类蛋白的帮助下，单链分支侵入到双链分子中。

4) 环切割（loop cleavage）。在单链分支入侵后，被侵入的双链分子中的一条链由侵入者所代替，被替代的链在酶作用下除去，除去过程可能是起初断开，随后被内切核酸酶或外切核酸酶或两种酶作用除去。

5) 同化（assimilation）。上述第4步中形成的缺口的3′-OH可被DNA聚合酶延长（与第2步中所述过程相同），并被连接酶连接到侵入的那条DNA链上，进而缺口（gap）扩大使侵入链与被侵入分子能更多地配对，在这一点形成对称中间体。配对是稳定的，但只存在一个异源双链区（指上面分子中），因为复制过程由于填补了缺口而引起下面分子中的一条链的移动和转移。在Holliday模型中，缺口通过上面分子中的一条链向下转移来填补，所以可以存在两个异源双链区。

6) 异构化。参与重组的两个DNA分子的每条链发生重排并且形成Holliday交叉。Holliday结构中的链交换变成新的骨架链，前边的骨架链变成新的Holliday交叉。

7) 支链迁移。Holliday交叉迁移（如图6-18中向右迁移）产生一个具有两个异源双链区的对称排列，这是由于在同化和支链迁移中碱基配对的移动。当解离发生时，切口并不是在典型的入侵链或最初断裂处产生，这点与Holliday模型不同。

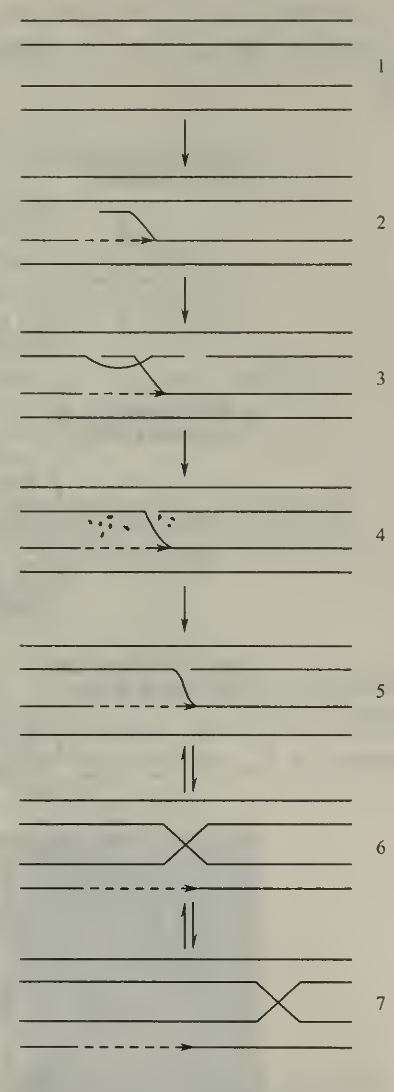


图6-17 不对称链转移模型

8) 解离（resolution）。分子分离与Holliday模型中一样，也存在选择南-北向或东-西向切割的问题。

图6-18可以用不对称链转移模型解释真菌中减数分裂重组的结果。图6-18（a）表明通过异构化而没有支链迁移形成的不对称中间体解离的产物。不对称链转移模型南-北切割侧旁标志不是重组体，而东-西切割产生重组体，两种情况下，两个分子中只有一个异源双链区，重组排列在左边可以解释不对称的5:3分离。图6-18（b）表明通过异构化和支链迁移所形成的对称中间体产物的交替配对。存在两个异源双链区的结构通过错配修复产生错误的4:4，5:3和6:2分离。不对

称链转移模型不包括对称中间体经过支链迁移而不经过异构化形成过程的可能性，在这种情况［图 6-18（c）］中，重组和非重组体的形成分别通过南-北和东-西切割过程。

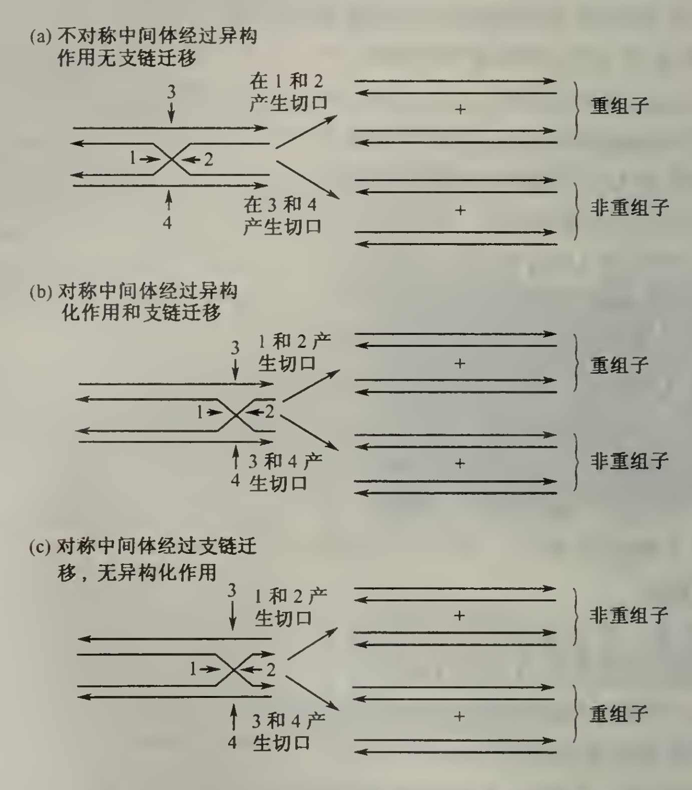


图 6-18　对称和不对称中间体变换产生不同产物

三、8 字形分子的解离

图 6-19 表示两个环状分子由于交叉形成的 8 字形分子可以通过三种方式切割。如果将 8 字分子的上边环转 180°则形成一个开环结构，首先假设这个结构易发生断裂，且断裂只发生在一个单链区域，如图中箭头所示的四个位点，断裂后将产生三种类型的分子，这三种形式在不同的系统中将起很不同的作用。例如，如果参与重组的是细菌或真菌的 DNA 分子，不管是二聚体或滚环式复制都必须加工成单个基因组长度的分子，但如果是在 λ 噬菌体系统中，产生单体分子将会导致死亡，因为 λ 噬菌体在包装时必须存在两个 COS 位点，只有二聚体或滚环（继续复制）复制才能产生重组噬菌体。关于滚环形式介导重组还无任何证据，但 $RecA^+$ 的细菌中的小质粒常发生二聚化。

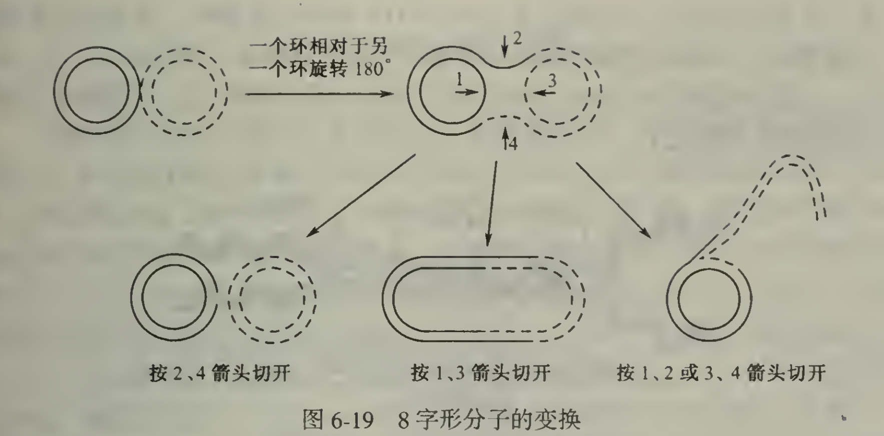


图 6-19　8 字形分子的变换

第五节　RecA 和 RecBCD 蛋白在重组中的作用

一、RecA 蛋白

在体外 RecA 蛋白能使 Holliday 结构产生，RecA 蛋白启动单链转移到双链 DNA 分子中，紧接着发生交换。图 6-20 表示在体外 RecA 蛋白使两个双螺旋（其中之一在一端产生一个切口）完整的重组接合，当退火时由 RecA 蛋白驱使到

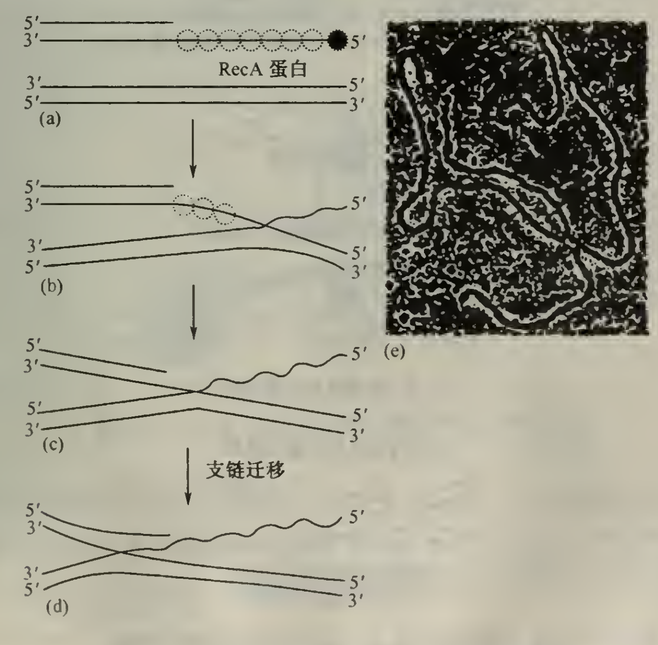


图 6-20　RecA 蛋白在体外介导的重组反应模型

gap 处［图 6-20（c)］，当上边分子 gap（切口）的左边松旋时，结合处将扩散到整个双螺旋区，同时在交换的右边形成第二个杂交双螺旋［图 6-20（d)］。这就说明 RecA 蛋白可能引导交换位点的左边区域形成四链螺旋区，促进整个双螺旋区的链的交换（图 6-21)。

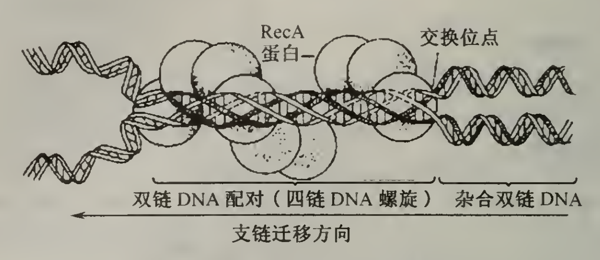


图 6-21　重组中的 DNA 四链螺旋

图 6-22 是一个精细的反应模型，可用来解释当 RecA 蛋白结合到 DNA 上一远离末端的切口时，在细胞内如何启动重组。RecA 蛋白作用的靶 DNA 双螺旋中必须产生一个切口，提供一个自由末端并由此链互相缠绕形成双螺旋。现已发现一些参与重组的蛋白质的内切核酸酶活性。

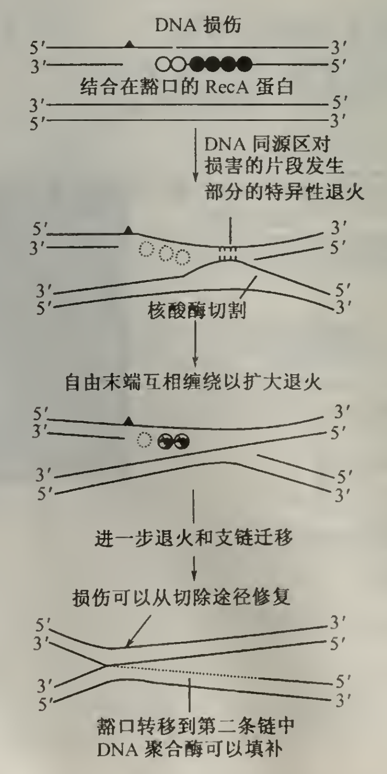


图 6-22　从一个豁口处起始重组的类似途径

二、RecBCD 酶

大肠杆菌 RecBCD 酶具有多种活性：解旋酶活性、外切核酸酶活性和 ATP 酶活性。RecBCD 酶复合物由 *E.coli recB*、*recC*、*recD* 基因编码的蛋白组成，专一地识别 DNA 上的断裂处，这种断裂可由细菌接合转移、噬菌体的转导以及射线照射和某些化学物质处理后产生。在体外系统中，纯化的 RecBCD 酶可识别并结合于 DNA 的平齐末端（等于双链 DNA 的断裂处）后沿 DNA 移动，在移动的同时依赖其解旋酶活性解开双螺旋（图 6-23），它的 3′→5′外切核酸酶活性和 5′→3′外切核酸酶活性降解两个单链，当到达 CHI 位点时，3′→5′外切核酸酶活性降低，解旋继续进行并且 5′→3′外切核酸酶活性增强而产生单链 3′末端，此时 RecD 解离出来，RecBC 继续前进，产生的单链 3′末端由 RecA 蛋白结合覆盖形成重组末端（recombinogenic end）。

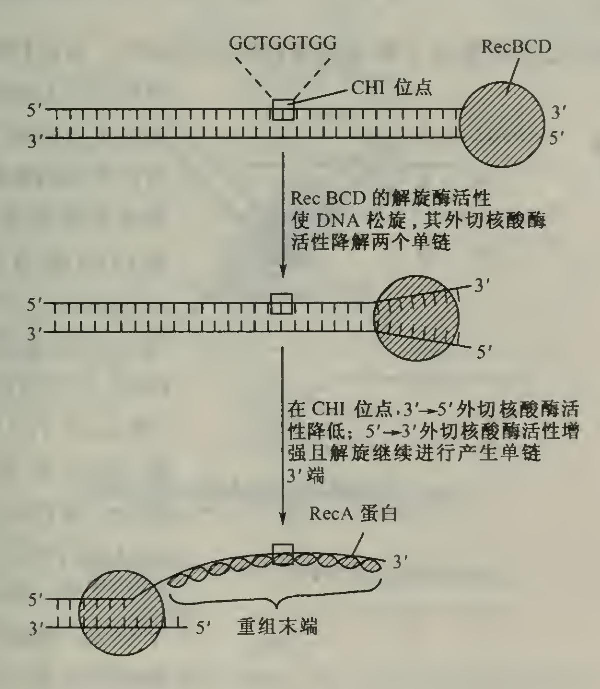


图 6-23 由 *E.coli* Rec BCD 酶介导的重组的起始

在研究 λ 噬菌体突变株时发现 λ DNA 上只改变一个碱基对就可以创造出一个激发重组的位点，这个位点称为 CHI 位点。CHI 位点的一致序列为不对称的 8bp 序列： 5′GCTGGTGG3′ / 3′CGACCACC5′ 。在大肠杆菌染色体上约每 5～10kb 就有一个 CHI 序列，这为 RecBCD 酶提供了许多机会。在正常细胞里这种序列对 RecBCD 酶通

常是不合用的，因为这些DNA没有自由末端，但对于细菌接合转移，噬菌体转导中的重组特别重要。事实上，正是RecBCD酶在λ噬菌体重组中的功能才导致了CHI位点的发现。

三、参与同源重组的其他蛋白质

1. Ruv蛋白

Ruv蛋白参与同源重组中支链迁移和Holliday结构的解离。虽然Holliday结构的形成依赖于RecA蛋白，但结构的维持却不需要RecA蛋白。除去RecA蛋白后留下的稳定Holliday结构可用于支链迁移和解离的研究。Holliday结构中交叉点的迁移可由大肠杆菌Ruv A和Ruv B蛋白有效地催化（图6-24）。进一步的研究显示，Ruv A专一地识别Holliday交叉点，Ruv B的解旋酶活性对支链迁移的发动是必需的。

近期的研究已经清楚地了解了这两种蛋白的作用。有活性的Ruv A以四聚体形式与Holliday交叉点结合，使交叉点形成非折叠状态的四方形平面构象并保持4条DNA单链片段处于分离态。Ruv A蛋白的结合诱导两个环状（ring-like）六聚体Ruv B蛋白与Ruv A复合物周围的双链DNA上对应的两个位点结合［图6-25（a)］，水解ATP供能，Ruv B六聚体环（ring）起分子泵作用，推动两个双链DNA分子进入Ruv A复合物，分开4条DNA链，然后将两个异源双螺旋区从Ruv A复合物中挤压出去，接着就是支链迁移。两个Ruv C内切核酸酶结合到Ruv A /Ruv B复合物上并在相对180°的两个位点切开DNA,经连接酶连接后产生重组（或非重组）的含异源双链区的DNA分子［图6-25（b)］。

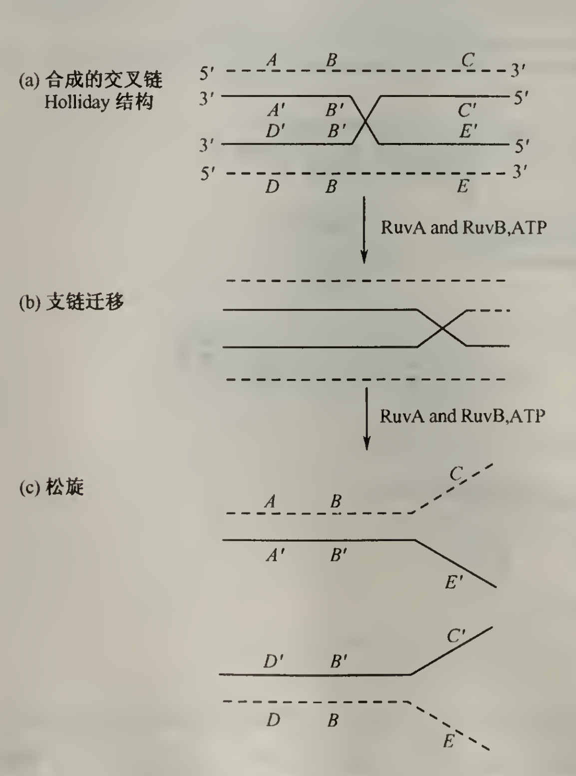


图6-24　由*E.coli*蛋白RuvA和RuvB催化的支链迁移

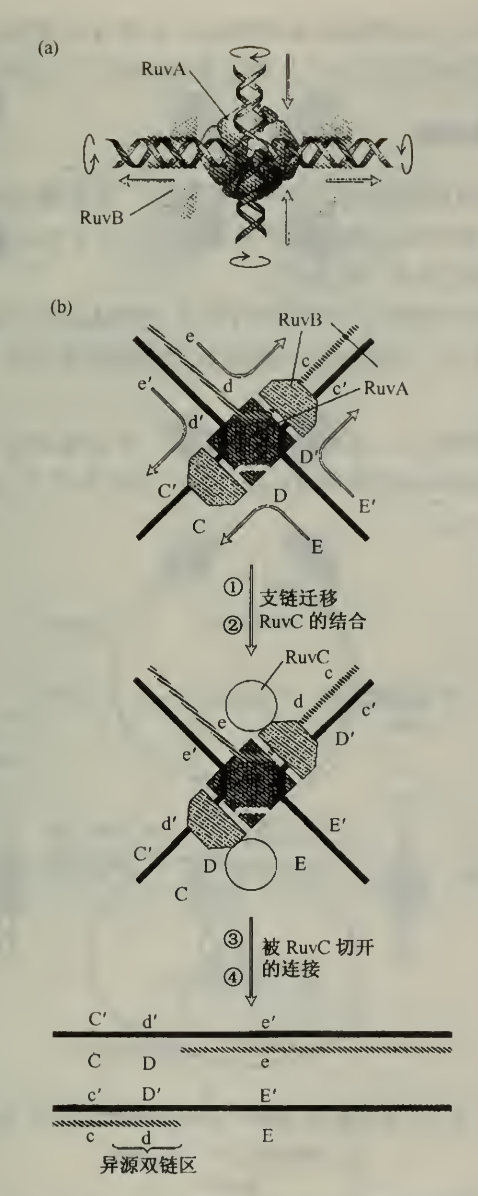


图 6-25 *E.coli* 蛋白在支链迁移和 Holliday 连接处解离中的作用

2. 真核同源重组酶

所有真核细胞包括人类细胞产生的重组相关蛋白在结构和功能上都与大肠杆菌中的 RecA、RecBCD、Ruv A、Ruv B、Ruv C 类似，例如在人和酵母中的 RAD51 蛋白，它在催化同源 DNA 片段配对、介导单链插入以及顺序的同源方面

都与 RecA 类似，而且已在细菌和真核细胞中发现有同源蛋白。由此可见在所有类型的细胞里参与同源重组的分子及同源重组机制几乎是类似的。

3．Cre 蛋白和其他重组酶

Cre 蛋白和其他重组酶催化位点专一性重组。重组酶识别位点专一性重组中的两个 DNA 分子中相对短的独特顺序并催化两个分子的连接。在原核和真核细胞中已经发现了几种位点专一性重组。

专一位点重组研究得最为充分的例子是 λ 噬菌体在宿主染色体特殊位点的整合。整合可在体外进行，只需要整合酶和宿主编码的整合因子。整合酶也催化原噬菌体的割离。

对 P1 噬菌体编码的 Cre 蛋白催化位点专一性重组的机制了解得最为清楚。当 P1 复制时，产生很长的多聚体 DNA，通过多聚分子上 loxP 位点的重组产生 P1 单体分子（图 6-26）。

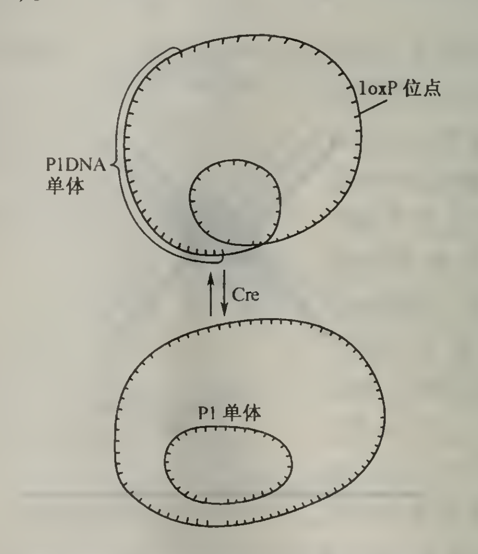


图 6-26　Cre 蛋白，由噬菌 P1 编码，催化多聚体环状 P1 DNA 上 loxP 位点之间的位点专一性重组

Cre 蛋白与 λ 的整合酶在结构与功能上都非常类似，甚至连 P1 单体分子从多聚体割离的机制也与 λ 原噬菌体的割离类似。

Cre 蛋白催化的重组见图 6-27，这个机制中涉及连续的两个中间体，在中间体中 Cre 和 DNA 通过磷酸酪氨酸键共价连接，类似于 Topo II 与 DNA 之间的连接（图 6-28），第一个 Cre－DNA 中间体反应产生类似 Holliday 结构，第二个反应产生重组体双链 DNA 产物。通过研究突变的 DNA 分子或是在 loxP 位点断裂的分子，研究者已能阻断 Cre 蛋白在中间体中的几步催化反应（图 6-29）。

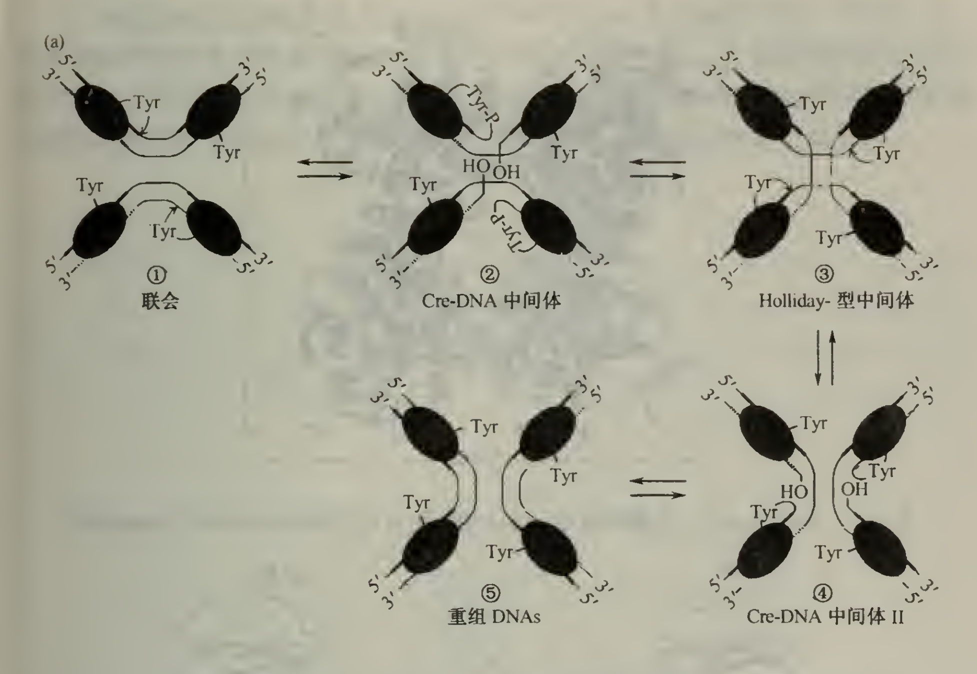


图 6-27 Cre-loxP 位点专一性重组的机制

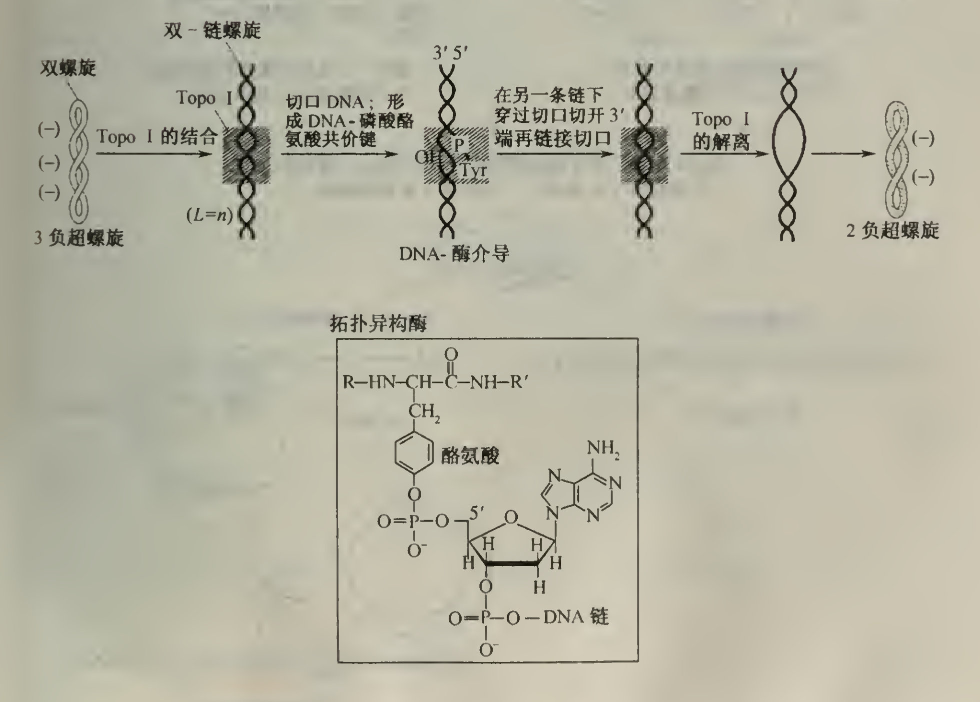


图 6-28 *E.coli* 拓扑异构酶 I 的活性

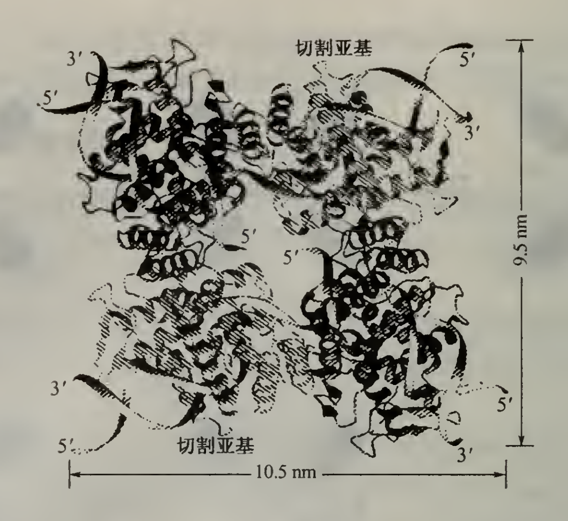


图 6-29　根据位点专一性重组中 Cre-DNA 中间体 II 的 X 射线结晶学的 Ribbon 模型

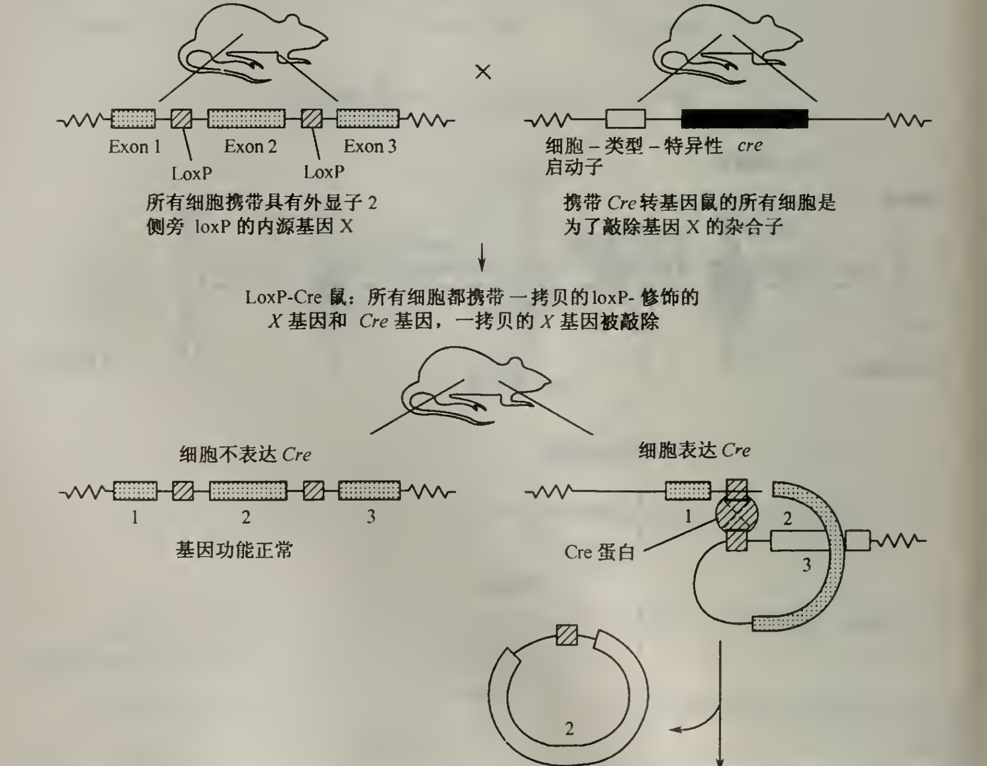


图 6-30　利用 loxP-Cre 重组系统进行细胞-类型-特异性基因剔除

P1 噬菌体的重组酶系统为发育生物学研究中的基因剔除提供了有用工具。经典的基因剔除老鼠是通过图 6-30 中的程序建立的。P1-loxP 重组酶系统已能使研究者建立仅在某一特别组织中缺失特殊目的基因的动物，这样的组织特异性基因剔除老鼠使研究者能对成年动物某一组织中任一基因的功能进行研究。

第七章　转　　录

基因表达的完成必须通过遗传信息从 DNA 到 RNA，然后从 RNA 到蛋白质这样的过程。RNA 分子以 DNA 分子的一条链为模板，在依赖于 DNA 的 RNA 聚合酶（DNA-dependent RNApolymerase）催化下合成（图 7-1）。依赖于 DNA 的 RNA 聚合酶也简称为 RNA 聚合酶（RNA polymerase）。RNA 分子合成的起始、延长、终止的过程叫转录。

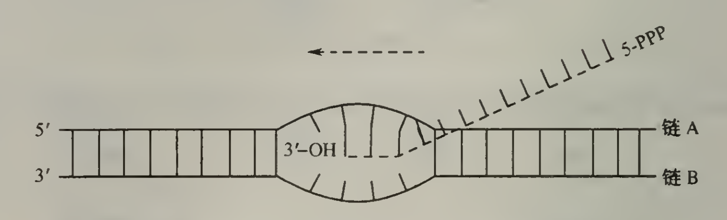


图 7-1　RNA 只从 DNA 分子的一条链拷贝

转录必须从两个方面进行讨论，第一是转录的酶学，第二是决定转录在 DNA 分子上的什么位点开始和停止的信号。

第一节　RNA 的酶促合成

一、RNA 酶促合成的基本特征

1）RNA 合成的底物是 5′-三磷酸核苷酸（NTP），包括 ATP、GTP、CTP 和 UTP，每个 NTP 的 3 位和 2 位碳原子上各有一个—OH。

2）在聚合酶作用下，一个 NTP 的 3′-OH 和另一个 NTP 的 5′-P 反应，去掉焦磷酸，形成磷酸酯键。

3）RNA 的碱基顺序由其模板 DNA 的顺序决定，NTP 依靠与 DNA 上的碱基配对的亲和力被选择，这一点与 DNA 复制相同，只是 dTTP 由 UTP 代替。

4）在开始转录的双链 DNA 分子的任何一个特定区域都是以单链为模板，如图 7-1。

5）RNA 合成的方向是从 5′～3′，单核苷酸只加到 3′-OH 上，新生 RNA 链与其模板 DNA 链呈反向平行。

6）RNA 的合成中不需要引物（这一点与 DNA 的合成不同）。

7) 只有 5′-三磷酸核苷酸掺入到 RNA 的合成中，所以起始转录处的第一个核苷酸 5′末端具有三磷酸结构，其 3′-OH 是下一个核苷酸的接触点，因而这样合成的 RNA 分子的 5′末端具有三磷酸。

二、大肠杆菌 RNA 聚合酶

大肠杆菌 RNA 聚合酶（*E.coli* RNApolymerase）由多亚基组成，全酶分子质量为 465 000Da（图 7-2）。通过对大肠杆菌突变株的研究，已经证明组成全酶的五种亚基由四个基因编码。*RpoA* 编码 α 亚基，*rpoB* 编码 β 亚基，*rpoC* 编码 β′亚基，*rpoD* 编码 σ 亚基。

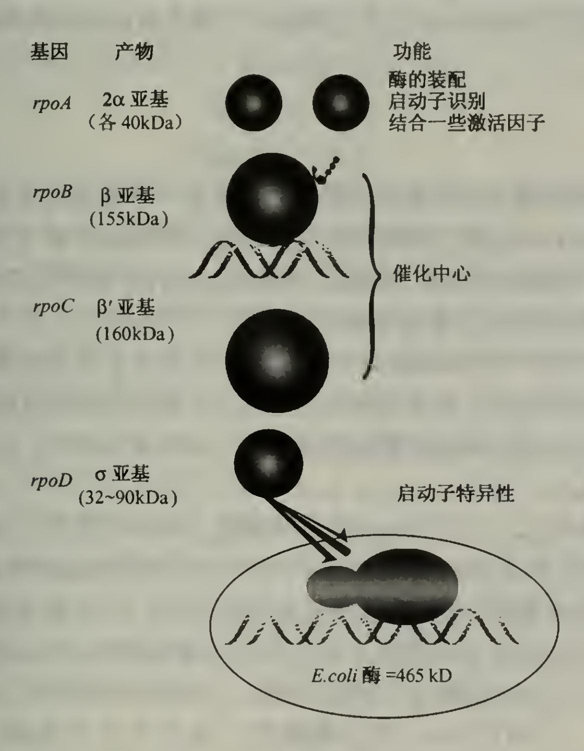


图 7-2　真细菌 RNA 聚合酶具有 4 类亚基

α，β 和 β′的大小在不同的细菌中相当恒定，σ 亚基的大小在不同的细菌中变化较大。

σ 亚基很容易从全酶上掉下来，σ 亚基在酶作用的某个阶段掉下来之后，剩下的 $\alpha_2\beta\beta'$，称为核心酶（core enzyme），而 $\alpha_2\beta\beta'\sigma$ 称为全酶。σ 是识别因子。在 *E.coli* 中，rRNA、tRNA 和 mRNA 均由同一种酶即 RNA 聚合酶（$\alpha_2\beta\beta'\sigma$）催化合成。

在 RNA 聚合酶的各亚基中，α、β 和 β′亚基对酶活性的重组是必需的，核心

酶具催化活性，其作用是识别 DNA 分子上 RNA 合成的起始信号。但 σ 亚基不能单独与 DNA 结合，它结合到核心酶后可能引起酶构型的变化，因而改变核心酶与 DNA 结合的特性。根据单个亚基或单个亚基聚集体的功能特征研究以及亚基的特异探针和变异体混合重组的研究，β 亚基是仅有的可以单独与 DNA 结合的亚基，它参与 RNA 聚合酶与模板的反应，同时 β 亚基也参与核心酶和 σ 亚基的结合以及转录的终止。β 亚基具有利福霉素和链霉素的结合点，也与 σ 和 α 亚基的结合有关，它参与 RNA 链的引发和延伸，催化底物形成磷酸二酯键。β（或 β′）亚基也可与终止因子 ρ 发生直接反应。α 亚基具有与 β 亚基的结合位点，参与特定的基因表达，可能与酶和 DNA 上的启动区域的反应有关。

除了四个多肽亚基外，大肠杆菌 RNA 聚合酶还含有 2 个 Zn 原子与 β 亚基相连接。无论在体外或体内，RNA 聚合酶通过下列几步反应聚集而成：

$$\alpha_2 + \beta — \alpha_2\beta$$

$$\alpha_2\beta + \beta' — \alpha_2\beta\beta'$$

$$\alpha_2\beta\beta' + \sigma — \alpha_2\beta\beta'\sigma$$

细菌 RNA 聚合酶在体外低离子强度下会发生一种可逆的凝聚。例如大肠杆菌全酶在 0.1mol/L 以上离子强度时以单体存在，在较低的离子强度下则形成二聚体，核心酶在较低离子强度下则是以六聚体以上的多聚体形式存在。

所有 RNA 聚合酶除了都具有相同的功能，即合成 RNA 以外，还可能与转录调节有关。如枯草杆菌芽孢化是与 σ 因子的失去和 β 亚基的修饰联系在一起的，因而枯草杆菌的 RNA 聚合酶突变株可阻止该菌的芽孢化过程。

原核生物 RNA 聚合酶能催化所有类别 RNA 的生物合成，但其活性能被利福平等药物抑制。

RNA 聚合酶很大，它与 DNA 结合时覆盖 DNA 的范围大约是 40bp（base pair），这个范围是通过 RNA 聚合酶与 DNA 结合后再用核酸酶降解 DNA 的方法确定的。当用 DNA 酶对 DNA 进行降解处理时，由于 RNA 聚合酶的保护，结合有聚合酶的那段 DNA 就不被降解，对这部分不被降解的 DNA 进行估计，得出大约是 40bp。这个方法就是大家所知的 DNA 酶保护法（dnase protection method）。另外一个测定 RNA 聚合酶覆盖范围的方法是硫酸二甲酯保护法（dimethy sulfate protection method），这个方法主要是利用 DNA 被 RNA 聚合酶结合后受到保护而不再被甲基化。

三、RNA 聚合酶在 DNA 上的识别结合位点

1. 启动子

转录的第一步就是 RNA 聚合酶结合到 DNA 分子上，与 RNA 聚合酶结合并

在此起始转录的特殊位置叫启动子（promoter)。启动子是 20～200bp 的专一序列，在这段序列中，发生一系列的相互作用。启动子是在 Lac^- 突变株中首次发现的，这个突变不仅破坏了 *lac* 基因的活性，阻止 Lac RNA 的转录，而且不能互补，叫做启动子突变（promoter mutation)。

在启动子处发生一系列的相互作用。RNA 聚合酶必须识别一个特殊的 DNA 序列，结合到一个合适的构型中，打开 DNA 双螺旋从而进入其中对其碱基进行拷贝，接着启动 RNA 的合成。这些事件的发生由 DNA 的碱基序列、RNA 聚合酶的 σ 亚基（没有 σ 亚基，启动子不被识别）共同引导，有些启动子还需辅助蛋白（auxiliary protein）引导。已知这个过程可归纳为三步：①RNA 聚合酶结合一个识别位点。②聚合酶移动到起始位点。③形成一个开放启动子复合物（open-promoter complex)。分离不同基因的启动子并测定它们的碱基序列，发现不同启动子的启动区非编码链有 7 个碱基是很一致的（TATAATG)。

2．-10 区

E.coli 和 *E.coil* 噬菌体的几个基因的启动子序列中，在左边距 mRNA 第一个开始转录的碱基 5～10 个碱基的地方有一段顺序叫 Pribnow box 即 -10 区，典型的 Pribnow box 序列为 TATAATG。在 Pribnow box 中，第六个碱基全是 T，即 Pribnow box 的第六个碱基是绝对保守的。已测知 35～36 个 *E.coli* 的 Pribnow box，其前两个碱基全部是 TA，突变体可以变成 TG、CA、GA 或 TC，但均保留了 TA 两个碱基中的一个。目前认为 Pribnow box 决定转录的方向，所以转录从左到右。在 Pribnow box 区，DNA 双螺旋解开并与 RNA 聚合酶形成 open-promoter complex。转录的方向称为下游（down stream)，所有上游（upstream）的碱基是不被转录的，不转录的碱基以负数表示。Pribnow box 在 -13～-4 之间。

图 7-3 是 *E.coli lac* 基因启动子的六个突变（箭头下的是突变的)，有两个突变在 Pribnow box 中，另一个在其旁边，Pribnow box 中的突变或者破坏转录的启动或者改变转录起始的必要条件，可见 Pribnow box 的重要。Pribnow box 的碱基顺序基本上是恒定的，测定了 100 个启动子，其结果如下：

T	A	T	A	A	T
89	89	50	65	65	100

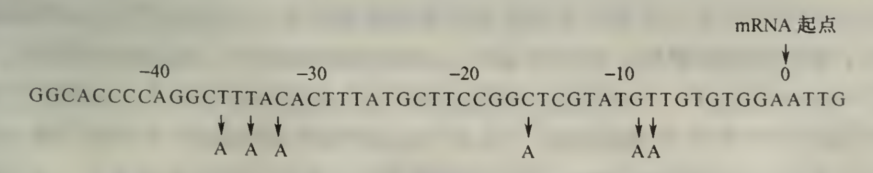


图 7-3 *lac* 启动子的 -10 区和 -35 区，下边标出突变位点和替换或缺失的碱基

3. －35 区

－35 区位于 Pribnow box 左侧，是启动子中另一个重要的区域，在－35 区也存在与 Pribnow box 类似的一个碱基序列，这个共同序列是 TTGACA。

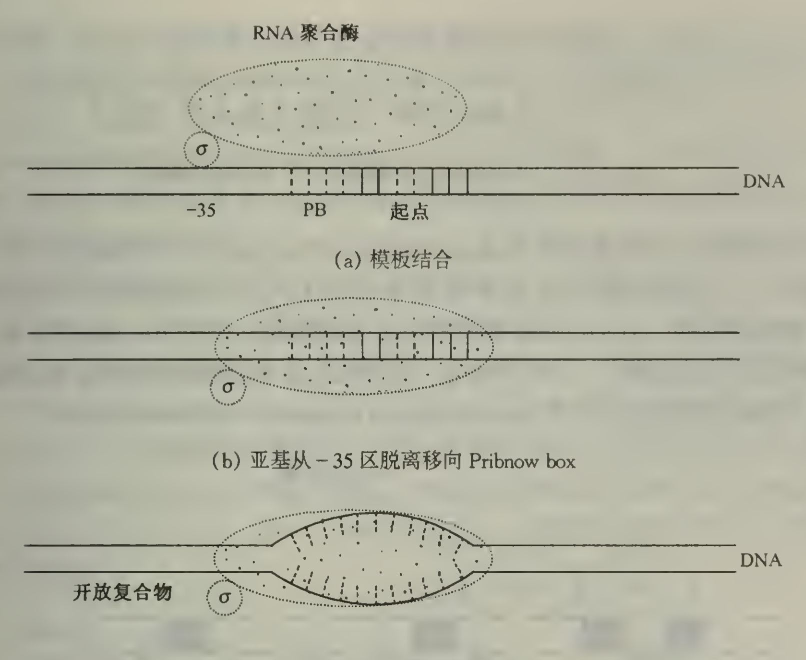
测定了 100 个 *E. coli* 的不同启动子的结果如下：

T	T	G	A	C	A
82	84	78	65	54	45

1975 年 Pribnow 比较了几种大肠杆菌噬菌体和大肠杆菌的启动子序列并发现了－10 区存在一致序列之后，Schaller 等人又发现，若把被全酶保护的 DNA 片段提纯，RNA 聚合酶全酶不能再和此片段结合，或者不能再选择正确的起始位点，因而推测，在酶保护区外一定还有一个序列对于酶识别启动子是必需的。对 λ 噬菌体和包括真核病毒 SV40 的研究结果表明，启动子区域的 DNA 序列必须从 mRNA 的合成起点向左延伸到 35 个碱基对以上，而乳糖操纵子的启动区域必须从 RNA 合成起点向左延伸到 60 个碱基对以上。λ 启动子和 SV40 启动子的－35 区都有共同序列，其中含有 *Hin*d 限制性内切核酸酶的位点，用 *Hin*dIII 处理不同的启动子后，启动子便失去活性。在－35 区的同源序列中，TTC 是很保守的。目前认为 RNA 聚合酶首先识别并结合于－35 区，然后移向 Pribnow box，至于是如何移动的目前还不很清楚，普遍认为 RNA 聚合酶结合后沿着 DNA 滑动。另一个可能性是 RNA 聚合酶的亚基首先与高度专一性的－35 区左边的识别位点相结合，然后，由于 RNA 聚合酶体积巨大而能进入到 Pribnow box 区，如图 7-4 所示。RNA 聚合酶一旦进入到 Pribnow box，就从左边的识别位点解离出来，和 Pribnow box 紧密结合，形成一个 open-promoter complex。

这个 open-promoter complex 是一个高度稳定的复合物，并且起着链启动的中间作用。在这个复合物中，距 Pribnow box 左边大约 10 个碱基对的位置 DNA 解旋，并且延伸到第一个被转录的碱基。这种解旋是核苷酸进入与 DNA 进行碱基配对所必需的。由于 Pribnow box 的碱基组成中富含 A、T，所以 Pribnow box 处 DNA 易变性。总而言之，RNA 聚合酶本身可诱导 Pribnow box 到 open-promoter complex 的结构变化。

我们把启动子分为强启动子和弱启动子两类来进行讨论。弱启动子被 RNA 聚合酶识别的效率较低，启动合成 mRNA 的数目也比强启动子要少得多，因而翻译出的蛋白质也少得多。强弱启动子是根据测定其合成蛋白质分子的多少来确定的。根据目前的研究，强弱启动子之间的差别就在于－35 区的序列不同。此外，Prihnow box 和－35 区之间的距离和启动子的强弱也有关系。在测定了 100 个不同的启动子的序列后发现－10 区和－35 区之间的距离 90% 以上是 16～18bp，也有少数是 15～20bp。虽然这段序列在启动子中可能并不重要，但 RNA 聚合酶的几何学却可能严格要求保持两个位点之间适当的距离，－10 区和－35



(a) 模板结合

(b) 亚基从 – 35 区脱离移向 Pribnow box

(c) 开放启动子复合物的建立

图 7-4　RNA 聚合酶和启动子形成开放启动子复合物

区之间相距 17bp 时启动子是最强的。

如果突变发生在启动子上，那么使 – 10 区和 – 35 区越远离标准顺序的突变，转录的 mRNA 越少，这种突变称之为 down mutation；越接近标准的顺序的突变，转录的 mRNA 越多，这种突变称为 up mutation。

4．CAP 位点

有些启动子完全缺少 – 35 区的同源序列，例如 λ pre、gal P 和 ara BAD 启动子，这些启动子只有在一些正效应子分子存在时才有活性。例如 λ pre 只有在 λcII 蛋白存在时才有活性。*lac* 启动子有 – 35 区序列，但在 cAMP 受体蛋白 CAP 不存在时它不能和 RNA 聚合酶结合。*lac* 启动子中有两个 cAMP-CAP 结合位点，一个位点在 – 70 到 – 50（site I），另一个在 – 50 到 – 40（siteII）。位点 I 有反向重复序列，如图 7-5 所示，这似乎是最强结合位点的特征。

位点 II 本身是一个很弱的结合位点，但是当 CAP-cAMP 复合物结合于位点 I 时，CAP-cAMP 复合物对位点 II 的结合活性大大提高。一旦位点 II 被占据，RNA 聚合酶即与 – 35 区序列紧紧结合，接着与 – 10 区结合。CAP 激发 RNA 聚合酶与 – 35 区的结合是一种重要的调节作用，它对 Pribnow box 的功能发挥也具有重要意义。当无 CAP 时，RNA 聚合酶也能和 Pribnow box 结合，但是形成的

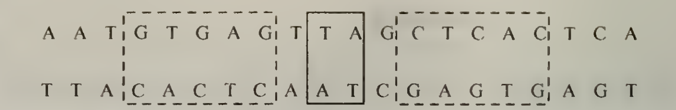


图 7-5　CAP 结合部位的碱基顺序，表现出对称性

是一个关闭的启动子复合物（closed-promoter complex），不能使起始区 DNA 双链解开。*lac* 启动子的 − 10 区序列是 TATG TTG，下边画线的 G 可能增加了 DNA 的热稳定性，所以 RNA 聚合酶不能诱导解链，而 CAP-cAMP 的结合可能使紧邻位点 I 的富含 G、C 区不稳定，在 Pribnow box 处降低 RNA 聚合酶的熔链温度，因而很容易建立一个 open-promoter complex，如图 7-6 所示。

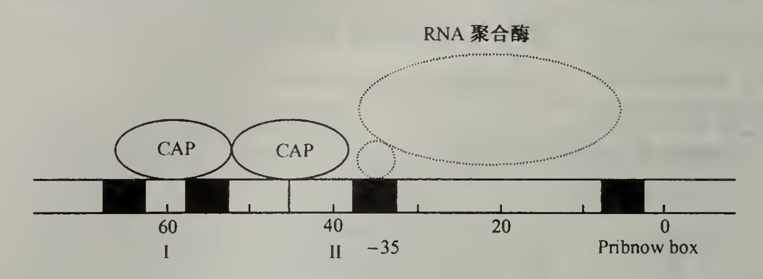


图 7-6　在 *lac* 启动子上有两个 CAP 位点和 RNA 聚合酶位点，I 和 II 是这两个位点

λ 噬菌体 UV5 突变株在 *lac* 启动子的 Pribnow box 中有两个碱基发生了改变，G-A 和 T-A，即由原来的序列 TATGTTG 变成了 TATAATG，仅两个碱基的改变就使得这种突变株的启动子在形成 open-promoter complex 时不需要 cAMP-CAP 的辅助。这是因为 GC 碱基对的丢失降低了整个序列的热稳定性，在不存在 CAP 的情况下 RNA 聚合酶也能解开 DNA 螺旋。有些有 − 35 区序列的启动子需要其他因子作用才能与 RNA 聚合酶结合，这些因子可能促进 RNA 聚合酶的结合。

四、转录的起始

关于原核生物 RNA 转录的起始过程，目前所了解的大致可概括如下：①RNA聚合酶核心酶在 σ 亚基参与下与 DNA 分子结合，形成非专一复合物，这样的复合物是很不稳定的；②起始识别：RNA 聚合酶全酶与启动子结合形成封闭的启动子复合物（closed-promoter conplex），这时酶与 DNA 的外部结合，识别部位大约在启动子的 − 35 处；③活化：形成开放的启动子复合物（open-promoter complex）时，RNA 聚合酶在 − 10 区与启动子紧密结合，并在解开 DNA 双链后

识别有义链。由于－10 区富含 A、T，故有利于 DNA 解链；④形成三元起始复合物：当 RNA 聚合酶移动到转录起始点时加入第一个三磷酸核苷酸得到三元复合物（启动子、全酶、三磷酸核苷酸），即开始转录。

RNA 链合成的起始是从形成开放启动子复合物开始的，一旦该复合物形成，RNA 聚合酶就开始准备合成 RNA。RNA 聚合酶上含有两个核苷酸结合位点，分别叫做起始位点和延伸位点（initiation site and elongation site）。起始位点只结合嘌呤三磷酸核苷酸即 ATP 和 GTP，这两个嘌呤三磷酸核苷酸之一（常常是 ATP）是 RNA 链上的第一个核苷酸（有时也见到有 UTP，但是很少），因此 DNA 上第一个被转录的碱基常常是胸腺嘧啶。起始的三磷酸核苷酸首先结合到酶上，再在 open-promoter complex 处与 DNA 上互补的碱基形成氢键（图 7-7）。延长位点与另一个三磷酸核苷酸结合，这个核苷酸是通过其对 DNA 链上起始位点下一个碱基的配对能力被严格选择的。之后，起始位点和延长位点上结合的这两个核苷酸再结合到一起，第一个结合到 DNA 上的碱基从起始位点上释放出来，链起始完成。



图 7-7　RNA 合成的起始

当 RNA 聚合酶释放碱基并且沿 DNA 移动时，链的延伸即开始。

如何选择 DNA 上第一个被拷贝的碱基目前还不清楚。它常常是在保留有 T 的－10 区序列的第 6 到第 9 个碱基之间，但是目前既没有确切的位点，也没有明确的模式（图 7-8）。

启动子	序列
E. coli galP1	5′-GTATG-3′
E. coli araBAD	5′-GTATG-3′
噬菌体 φ*X*17 基因	5′-GTATG-3′
病毒 SV40	5′-GTATG-3′

图 7-8　四种不同的启动子中的 RNA 的第一个碱基处的模板顺序

五、RNA 链的延伸

核苷酸不断加到 RNA 链的 3′-OH 上，RNA 链就延长。在起始到延伸的转变过程中，DNA 分子和酶分子可能发生构象的变化以适应延伸过程的需要。当

酶从起始位点往下游移动时，解旋也随酶一起进行，而转录完成部分的 DNA 则重新形成完整的双螺旋，这表明 DNA 螺旋在酶作用下有一个可逆的过程，同时酶分子的构象也发生了变化。

在转录起始阶段，全酶与 DNA 分子要形成稳定的复合物，因而对结合部位的一级结构有严格的专一性要求；而在延伸阶段，为了能沿模板移动，RNA 聚合酶必须放松对 DNA 的结合，而且对拷贝每一个核苷酸应具有同等能力，而不应有所偏向。RNA 聚合酶与 DNA 模板相互作用的这些变化，使聚合酶在链延伸几个核苷酸（估计是 8 个）之后就将 σ 亚基从全酶上释放出来，这样链延伸就由核心酶催化。σ 亚基的存在与否，引起 β 与 β′ 亚基构象上的变化，即当 σ 亚基结合时，β 与 β′ 表现为有利于专一与 DNA 结合的构象，而 σ 亚基释放后留下的核心酶则与 DNA 不专一结合，失去识别能力或优先与任一核苷酸序列结合的能力，不能形成稳定的复合物，如图 7-9 所示。

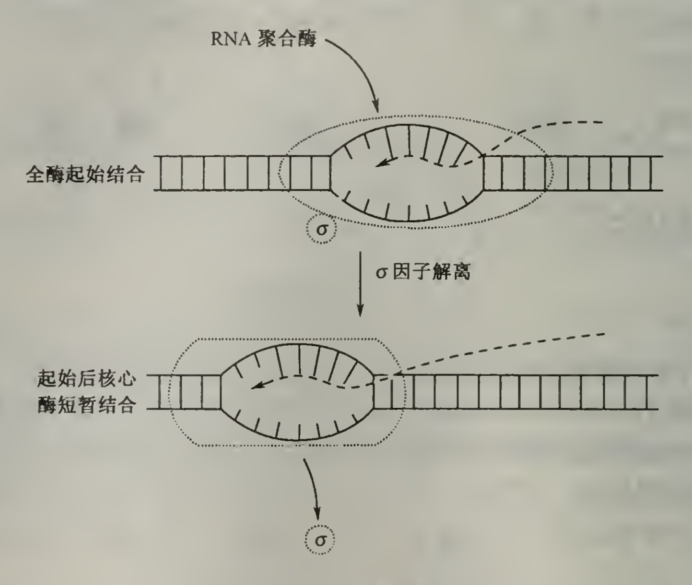


图 7-9　RNA 合成起始不久，RNA 聚合酶核心酶变构，σ 因子脱落

当核心酶沿 DNA 移动时，继续结合下一个能与 DNA 配对的核苷酸，并且解开前面的 DNA 螺旋。而原先解开的区域只有几个碱基对，恰恰就在 RNA 聚合酶的后边，其 DNA 螺旋又重新关闭，于是新合成的 RNA 就从重新关闭的 DNA 中释放出来。

用牛胰 RNA 酶（这种酶只作用于单股 RNA 但不切割 DNA-RNA 杂合分子）处理转录系统，结果表明大约有 10 个碱基被保护而不被酶切，这种保护作用究竟是由于 RNA 与 DNA 形成了氢键还是仅仅由于酶的大块空间的干扰，现在还不清楚。

链延伸的速度不是恒定的，在 DNA 的某些区域，转录速度减慢或停止，这种速度的降低叫 pause，pause 发生在富含 GC 区（图 7-10），在图 7-10 的这个序列里，由于 GC→AT 的变化，使得此突变株消除了 pause，因为 GC 变成 AT 后降低了氢键的稳定性。

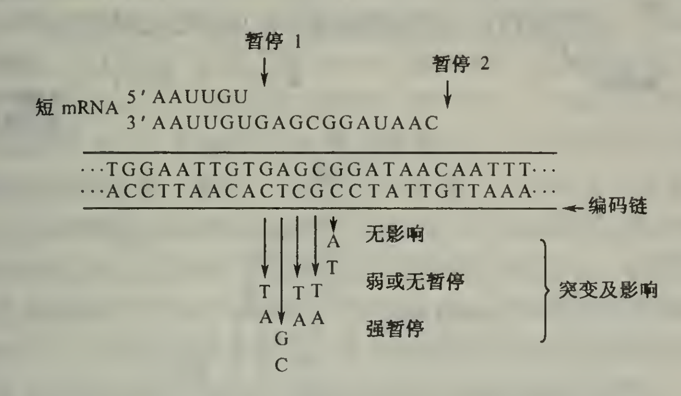


图 7-10 *E.coli lac* 系统的一个片段，有两个 pause 突变后可消除 pause（GT→AT）

六、RNA 链合成的终止和新合成 RNA 链的释放

RNA 合成的终止发生在 DNA 特殊的碱基序列中，对 20 个已测知的终止序列的研究表明，它们具有共同的特征（图 7-11），终止序列具有以下三个重要的区域：①具有一个反向重复序列，中间是不重复的片段，即假设 DNA 上一条链的序列是 ABCDEF—XYZ—F′E′D′C′B′A′，其中 A 与 A′、B 与 B′等是互补的（如色氨酸操纵子中前导区存在的终止区：AAAGGCTCCTTTTGGAGCCTTT），这样在转录时，DNA 链中可以通过链内碱基配对形成二元对称结构，因而转录的 RNA 在这部分可以形成茎-环结构，这种结构在转录终止中独立起作用，并且使得新合成的 RNA 链能抗 RNaseII 的降解（RNaseII 是一种胞内酶，对双股 RNA 无活性）。②第二个区域是在靠近推测的茎的环端（有时全部在茎中），是一个富含 GC 的区域。③第三个区域（有时不存在）是一个 AT 序列（可能在茎的开始），所以在 mRNA 中会转录产生一个 6～8 个尿嘧啶接着一个腺嘌呤的序列。

RNA 聚合酶移动至终止部位时转录产物仍然与模板链以氢键结合（图 7-12），此时 RNA 中茎的一半已经游离出模板，并通过 RNA 分子内相互作用而形成茎环结构。在 DNA 链的这个区域中二元对称结构的变异将对转录产物茎环的形成发生重要影响，而茎环的形成又能促使该处 DNA 链退火恢复螺旋结构，使茎环位于螺旋之外，这时 RNA 聚合酶就从“延伸”构象转变为“终止”构象。

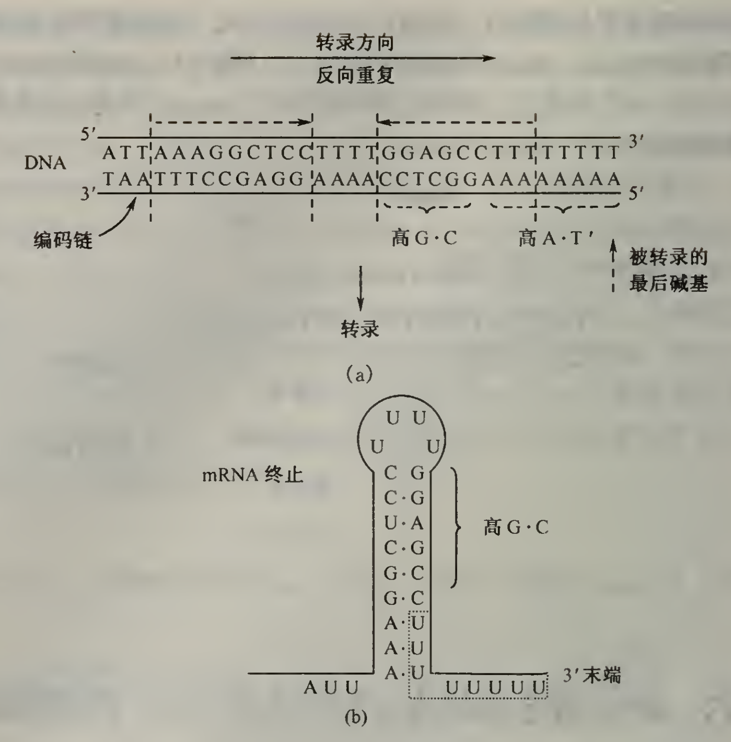


图 7-11　*E. coli* trp 操纵子终止区的 DNA 序列（a）和 mRNA 序列（b）

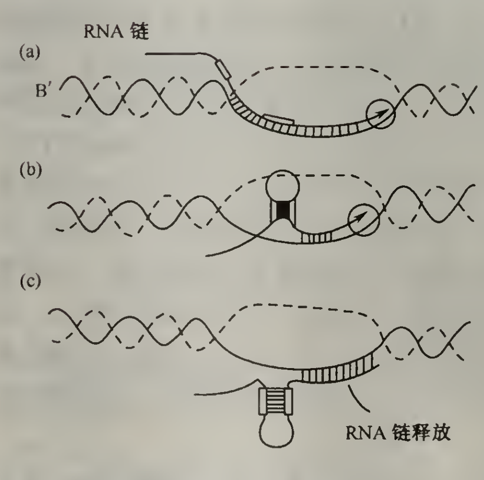


图 7-12　转录终止及链释放模型

此外，由于茎环后一般连着数个连续的 U，rU-dA 配对的 DNA-RNA 杂交体很不稳定从而释放出转录产物。

转录的终止有两种形式，一种是不依赖于终止蛋白 Rho（ρ）因子而靠 DNA 自身的序列终止转录，另一类是需要终止蛋白（ρ 因子）的作用终止转录。不同模板的终止子结构或其生成的 RNA 茎环的大小及其稳定性是不一样的，例如噬菌体 λ 6S 基因转录的 RNA 的茎是很长的，且富含 GC 对；而依赖于 ρ 因子的噬菌体 λ *Cro* 基因转录产物的末端茎最短，而且最不稳定。

最近对 trp 操纵子终止信号的研究表明，在操纵子最末一个结构基因之后存在两个终止信号，第一个信号具有典型的终止子结构特点，即在数个 U 之前存

在茎环结构，这被认为是一个体内转录终止信号。但如果体外加入 ρ 因子，则 RNA 聚合酶在第一个终止信号下游 250 个核苷酸处终止转录，该区并无典型的茎环结构特点，但 AU 丰富，因此可能存在另一种未被认识的终止机制。

在依赖 ρ 的终止子处结构不够稳定，RNA 聚合酶只是在此暂停但不终止，因而需要 ρ 的帮助才能停止转录。ρ 具有两种活性：①终止转录；②依赖于 RNA 的核苷三磷酸酶（NTPase）活性。同时许多试验表明，这两种活性为同一个蛋白质所具有，因此提出了一个转录终止模型（图 7-13）：开始 ρ 因子附着在新生 RNA 上，并随之沿 5′→3′方向朝 RNA 聚合酶移动，ATP 或其他三磷酸核苷酸水解为这种移动提供能量。最后，ρ 接触到在依赖于 ρ 的终止子处暂停的 RNA 聚合酶，与之反应使新生的 RNA 链释放出来。从这个模型可以清楚地看到，影响 ρ 与 RNA 的结合或 ρ 沿 RNA 移动的因素将抑制转录的终止。在原核生物中转录与翻译往往是偶联在一起的，翻译的同时进行意味着核糖体将遮盖新生 RNA，并阻止 ρ 的结合，从而抑制转录终止；而当核糖体与 RNA 脱离时出现裸露的 RNA，ρ 就发生作用而使转录终止，这就是操纵子末端终止的机制。

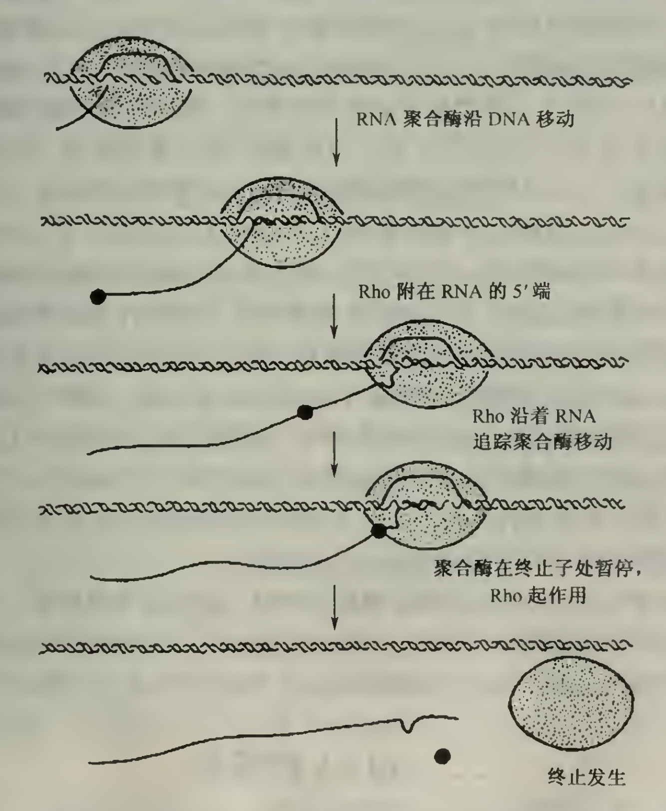


图 7-13　Rho 因子跟踪 RNA 聚合酶沿着 RNA 移动到 Rho 依赖的终止子时终止转录

第二节　RNA 分子的种类及转录后加工

RNA 分子主要分三种类型，即 mRNA（messenger RNA）、rRNA（ribosomal RNA）和 tRNA（tranfer RNA）。它们都是由 DNA 转录而成的，都叫 RNA，但在蛋白质合成中具有很不同的功能。

一、mRNA 的结构

细胞中每一条多肽链的氨基酸顺序其实是由 DNA 分子中的碱基顺序决定的，虽然氨基酸对 DNA 没有亲和性，但在 DNA 到蛋白质的转变过程中有一中间过程，在这中间过程中 DNA 上的遗传信息转录到信使 RNA（mRNA）分子上，再由 mRNA 作蛋白质翻译的直接模板。

与一个多肽链相对应的 DNA 片段加上启动子部位再加上终止子部位叫做一个顺反子，一次转录得到的 mRNA 可以含一个顺反子，也可以含多个顺反子，分别叫做单顺反子 mRNA（mono-cistron mRNA）和多顺反子 mRNA（poly-cistron mRNA）。实际上以多顺反子的方式转录是一种在调控上很经济的方法。

mRNA 的大小不一，小的可以是大约只编码 50 个氨基酸的 mRNA，也就是只有 150 个碱基；大一点的可以是能编码 300～600 个氨基酸的由 900～1800 个碱基组成的 mRNA。原核生物中的多顺反子 mRNA 比单顺反子 mRNA 更普遍，因此原核生物中一个典型的 mRNA 分子一般包含有 3000～8000 个碱基，少数巨大的分子，如组氨酸 mRNA 分子编码十个蛋白质，含有 12 000 个碱基。

一个顺反子除了翻译中的起始和终止序列外，在 mRNA 中还存在其他重要的区域。有些 mRNA 具有领头和尾部（leader and tailer），这两个区域是不编码蛋白质的，在某些情况中，leader 中含有一个调控区称为弱化子（attenuator），弱化子（有的人称为衰减子）决定蛋白质合成的速率。在 mRNA 的 5′-P 和 3′-OH 两端均发现有不翻译的序列，而且多顺反子 mRNA 分子中还可能含有几百个碱基长的顺反子间序列（intercistronic sequence）。

mRNA 一般以 DNA 的一条链为模板来转录，但并非都是这样，有许多噬菌体 mRNA 可以同时以两条链为模板从两个方向转录，例如噬菌体 λ 就是以两条链为模板进行转录，分别称为左向转录链（l）和右向转录（r）链。

二、mRNA 的寿命

原核生物 mRNA 的一个重要特点就是它的寿命比 tRNA 和 rRNA 的短，例如细菌，其 mRNA 的半衰期只有几分钟，lac mRNA 终止合成后 9min 就几乎全

部消失。mRNA 很容易从 5′-P 被降解，但有些 mRNA5′-P 末端被修饰而得到保护，所以不易被降解，例如真核生物中的血红蛋白 mRNA（hemoglobin mRNA）和蚕的蚕丝纤维蛋白 mRNA（silk fibroin mRNA）。

三、稳定的 RNA——核糖体 RNA 和转运 RNA

蛋白质合成除了需要 mRNA 作翻译的模板以外，还需要核糖体 RNA（rRNA）作翻译进行的场所，需要转运 RNA（tRNA）作为转运氨基酸的工具。rRNA 和 tRNA 在代谢中比 mRNA 稳定。

rRNA 和 tRNA 的转录都是从启动子起始到终止子结束来完成的，从这一点讲，rRNA 和 tRNA 的合成过程与 mRNA 没有什么不同，但 rRNA 和 tRNA 有以下三点特性与 mRNA 不同：①所有 RNA 初级转录产物的 5′末端均是 5′-三磷酸，而成熟 rRNA 和 tRNA 分子的 5′末端是 5′-单磷酸。②rRNA 和 tRNA 分子比其初级转录产物小很多。③所有 tRNA 除了含有 A、G、C 和 U 以外，还含有稀有碱基，这些稀有碱基在初级转录产物中是没有的。

rRNA 和 tRNA 分子的所有这些变化都是通过一个称为转录后修饰或转录后加工的过程来完成的。

四、tRNA 分子的加工

tRNA 分子的种类远比 rRNA 分子种类要多，在原核生物中 tRNA 分子有 30～40 种，真核生物中则更多，大约有 60 种以上。在原核生物中，有些 tRNA 基因是和 rRNA 基因连接在一起的，包括间隔区中的 tRNA 基因和末端的 tRNA 基因两种情况。例如大肠杆菌 K12 有七个 rRNA 基因簇，它的 16S rRNA 和 23S rRNA 基因之间都插有 tRNA 基因，其中三个编码 $tRNA^{Ile}$和 $tRNA^{Ala}_{IB}$，另外四个编码 $tRNA^{Glu}_{2}$；另外有几个 tRNA，如 $tRNA^{Asp}$、$tRNA^{Trp}$和 $tRNA^{Thr}$则由 rRNA 基因簇中远离启动子的末端序列转录。

此外，tRNA 基因大多以基因簇存在，构成多顺反子转录单位。大肠杆菌基因图上 tRNA 基因的已知位置已测定，约有 60 个 tRNA 基因，若考虑到摆动假说，这约为翻译 61 个密码子所需要基因的二倍。1981 年鉴定了一个大肠杆菌操纵子，其中含有七个 tRNA 基因：*Metm-Leu-Gln1-Gln1-Metm-Gln2-Gln2*（基因图的 15min 处），各 tRNA 基因间被二级结构隔开，推想为 RNA 加工酶的识别位点。

由于加工紧随着转录之后，转录还没有终止或刚开始不久加工就已经开始进行了，所以在野生型细菌中很难分离出大的转录产物。图 7-14 表示了双顺反子基因转录与加工的过程，一种为单顺反子产物，另一种是双顺反子产物，后者常

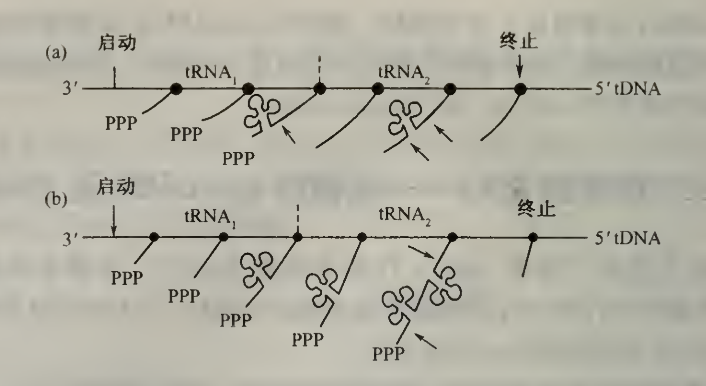


图 7-14　tRNA 的转录与加工关系示意图

在加工酶缺陷型菌中分离得到。两种产物都需要进一步加工才能形成成熟 tRNA 分子。

tRNA 前体的加工主要包括以下三个方面：①剪切、修剪和剪接；②如果需要，在 3′端添加 CCAOH；③核苷修饰。

所有这些反应都是在酶催化下进行的。总的来说，tRNA 加工所需要用的酶比 mRNA 或 rRNA 加工所需要的酶都要多，特别是多顺反子 tRNA 基因转录产物的加工更为复杂，往往需要多个酶的协同作用。

1. 剪切、修剪和剪接

tRNA 剪切酶是一类主要的加工酶，属内切核酸酶性质，它在 tRNA 前体两端的附加序列处水解前体 tRNA 以得到一定大小的 tRNA 分子。大肠杆菌 RNase P 特异地剪切几乎所有大肠杆菌及其噬菌体 tRNA 前体的 5′端附加序列，生成成熟的 tRNA 5′末端，因此这个 5′内切核酸酶也被叫做 tRNA 5′成熟酶。比较了几个 tRNA 前体的 RNase P 作用位点后发现，RNase P 的专一性与其剪切位点周围的一级结构以至二级结构并无直接关系。经研究发现 RNase P 的专一性可能与 tRNA 前体的三叶草特殊构型有关。另外，无 3′端 CCAOH 序列的 tRNA 前体不能作为 RNase P 的底物，所以 RNase P 的专一性也与 3′CCAOH 有关。

1978 年 Altman 等发现大肠杆菌 RNase P 是一种特异的分子，它由至多 23% 蛋白质和至少 77% RNA 组成，沉降密度为 1 .71g/cm^3（CsCl 中）和 1.55g/cm^3（Cs_2SO_4 中），它是一个典型的 RNA-蛋白质复合物。拆开蛋白质和 RNA，两者不具有活性，重组则活性恢复。

RNase F 是 3′内切核酸酶，它在距 3′端不太长的核苷酸序列处切开较大的前体。用于 rRNA 加工的 RNase III 和 RNase E 在 tRNA 加工中可能也起作用，但体内 tRNA 成熟并非绝对需要 RNase III，而且 RNase E 剪切也仅仅限于很少数

tRNA前体中，这两种酶的功能可能只是使大的新生转录产物变得小些，以便更有效地与 RNase P 作用。

为了使 tRNA 分子得到成熟的 3′端，还需要另一个外切核酸酶进行修剪，从 3′端逐个切去附加的序列以暴露出 tRNA 3′端。这个修剪酶是 RNase D，其分子质量为 38 000U，为单一亚基蛋白质。RNase D 不同于另一个 3′外切核酸酶 RNase II，它具更为严格的选择活性，因此是体内 3′tRNA 成熟酶。

所以，tRNA 前体中插入序列的除去是靠一个非常复杂的酶系来完成的。

2. 核苷修饰

成熟的 tRNA 分子中含有许多修饰成分，不同 tRNA 所含修饰成分的种类和数目都各不相同。虽然对于修饰成分在 RNA 中的生物功能仍了解很少，但对其生物合成过程已有不同程度的了解。核苷修饰问题是 RNA 特别是 tRNA 加工的主要内容。tRNA 的修饰是在多核苷酸链上进行的，由一类统称为修饰酶的酶类所催化。

无修饰成分的 tRNA（或前体）$\xrightarrow{\text{修饰酶}}$含修饰成分的 tRNA（或前体）。

tRNA 修饰酶具有高度的专一性，DNA、rRNA（可能有例外）多核苷酸或是单核苷酸、核苷都不能作为此类修饰酶的底物。由于 tRNA 中修饰成分繁多，因而修饰酶种类很多，一般来说，每种修饰核苷都有催化其合成的修饰酶。

tRNA 甲基转移酶（简称 tRNA 甲基化酶）是研究较早较详细的一种 tRNA 修饰酶，它不同于 mRNA 甲基化酶和 rRNA 甲基化酶。近年来由于分离纯化技术的提高，已经得到对碱基及其排列顺序高度专一的 tRNA 甲基化酶，如 tRNA（腺嘌呤-1)甲基化酶催化 tRNA 中特定位置的 A→mA，tRNA（鸟嘌呤-7)甲基化酶可催化 tRNA 中第 55 位的 G→m^7G，其他如催化 tRNA 中生成 m^1G、m^2G、m^5C 以及 rT 的酶也已高度纯化。tRNA 甲基化酶对碱基顺序具有严格的要求，如从酵母中分离的两个 tRNA（鸟嘌呤-1）甲基化酶，它们能分别甲基化 tRNA 中的 G_{19}和 G_{43}并分别生成 m^1G_{19}和 m^1G_{43}；从枯草杆菌分离的 tRNA（腺嘌呤-1）甲基化酶能催化 tRNA 中的 YAAGGCm_2^2G→YAm^1AGGCm_2^2G；而鼠肝中得到同一种酶则可催化 GTΨCGAAU→CTΨCGAm_1^1AU。

几乎所有 tRNA 甲基化酶的甲基供体均为 S-腺苷甲硫氨酸（SAM)，但也有报道在某些革兰氏阳性菌中分离出的所有 tRNA（尿嘧啶－5）甲基化酶（即 rT 甲基化酶）不是以 SAM 为供体，而是从5, 10-甲叉四氢叶酸获得甲基，同时以 $FADH_2$ 作还原剂。

广泛研究的 tRNA 修饰酶还有 tRNA 异戊烯转移酶（tRNA$\triangle^2$ 异戊烯合成酶）及 tRNA 鸟嘌呤转糖苷酶和 $tRNA_\Psi$ 合成酶。

下面具体介绍一个 tRNA 的加工过程。

大肠杆菌 $tRNA^{Tyr}$基因的每个基因重复单位有 350 个核苷酸：两个基因之间有 200 个核苷酸的间隔区，$tRNA^{Tyr}$本身是 85 个核苷酸，加上间隔顺序共 350 个核苷酸。

加工过程分以下三步。

(1) 3′-OH 末端的形成

这个过程包括一个内切核酸酶识别一个发夹环结构（图 7-15）和一个外切核酸酶识别 CCA 三碱基序列这样两个过程。转录产物在位点①被内切核酸酶降解之后，其上游的 7 个碱基被 RNase D 逐个去除（RNase D 是一种核酸外切核酸酶，作用在图 7-15 中②），然后 RNase D 再除去 CCA 下游的两个碱基 UC（图 7-15 中④），这两个碱基的切除过程是在 5′端加工之后进行的。在 *E.coli* 中曾分离到这两个碱基切除前的分子，称之为 pre-tRNA，*E.coli* 中 pre-tRNA 分子有以下结构：

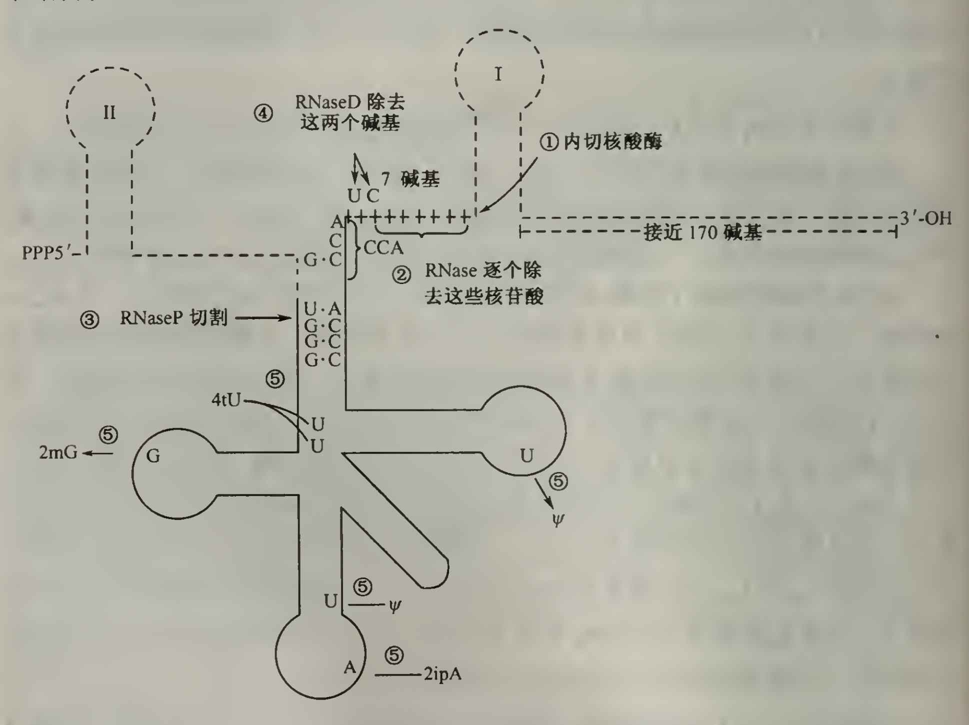


图 7-15 *E.coli* $tRNA^{Tyr}$转录子的加工过程

p-5′-（41bases）-tRNA-（2bases）-3′-OH，切除两个碱基后形成 CCA3′-OH 末端。

(2) 5′-P 末端的形成

5′-P 末端的形成在 RNase P 作用下完成，RNase P 切除图 7-15 中 5′端发夹

结构处的 41 个碱基，形成 5′-P 末端。

(3) 修饰碱基的产生

碱基修饰过程就是高度专一性的酶分别在其专一位点改变碱基，见图 7-15 中⑤，大肠杆菌 $tRNA_1^{Tyr}$ 中的两个 U 转变成假尿嘧啶（pseudouridine，Ψ），两个 U 转变成 4-硫尿嘧啶（4-thiouridine，4tU），一个鸟嘌呤变成 2′-O-甲基鸟嘌呤 (2′-O-methyl-guanosine，2 mG)，一个腺嘌呤变成异戊烯腺嘌呤（isopenteny-ladenosine，2 ipA）。

从图 7-15 中可以看出，tRNA 中不需要的碱基都在成熟 tRNA 的两边，而不在其中间，间隔区在两个 tRNA 之间。

五、大肠杆菌中 rRNA 的加工

E.coli 中的 rRNA 有三种：5S rRNA、16S rRNA 和 23S rRNA，分别有 120、1541 和 2904 个核苷酸。这些 rRNA 和另外几个 tRNA 全部包含在一个具有 5000 个核苷酸的转录子中，这些转录子的碱基序列很多已研究清楚了。图 7-16 是 *E.coli* 的一个转录子，图示从 5′末端到 3′末端。

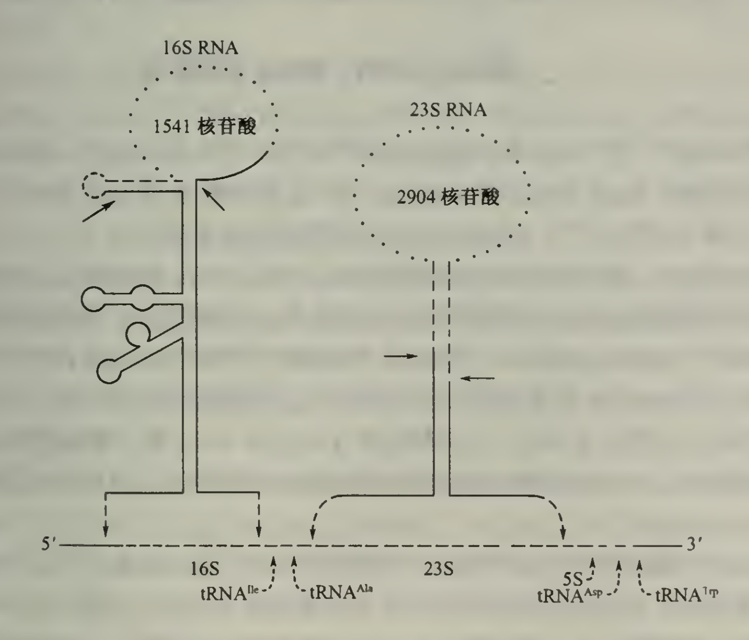


图 7-16　*E.coli* rRNA 转录子的加工

16S rRNA-$tRNA^{Ile}$-$tRNA^{Ala}$-23S rRNA-5S rRNA-$tRNA^{Asp}$-$tRNA^{Trp}$转录后的切割位点见图中的箭头。

当此巨大的转录子完成转录之后，一系列的酶就与之作用。虽然与之作用的酶系统还没有全部鉴定出来，但第一个酶显然是 RNase III，它是一种切割双链 RNA 的酶。首先 RNase III 在双链外切割，其他的酶接着完成 rRNA 的加工过程，RNase P 切割其中的 tRNA 部分。

在原核生物中还没有发现 mRNA 的加工，但有两个例外，一个是噬菌体 T7 中的 mRNA 被 RNase III 切割成 5 个不同的 mRNA 分子，另外，噬菌体 ϕ80 中一个 mRNA 是由 RNase P 加工而成的。

原核生物 RNA 的加工过程可以小结如下：①所有稳定的 RNA（tRNA 和 rRNA）转录后都需要加工，除噬菌体 T7 和 ϕ80 中的例外情况，mRNA 不需要加工。②大约有 10 个酶参与加工过程，一般内切核酸酶切割后产生 5′-P 和 3′-OH 基团。③3′-OH 末端一般由核酸外切核酸酶作用形成，5SrRNA 可能是例外。④许多加工酶对碱基不具有专一性，但要识别大的结构特征。⑤当不存在 RNase III 时，可能由 tRNA 的加工酶来代替。⑥稀有碱基是在转录以后通过修饰酶作用产生的，不是原来就有的。

第三节　真核生物中 RNA 的转录和加工

一、真核生物中 RNA 的转录

真核生物有三种 RNA 聚合酶，分别利用三类不同的启动子。RNA 聚合酶 I 转录产生核糖体 RNA 和大多数 snRNA，RNA 聚合酶 II 转录产生 mRNA 前体，RNA 聚合酶 III 转录产生 tRNA、5S RNA 和部分 sn RNA。

原核细胞中，除某些启动子需要辅助因子外，RNA 聚合酶可以单独起始转录。而真核细胞的 RNA 聚合酶不能单独起作用，它需要负责识别启动子元件序列的辅助因子。RNA 聚合酶 I 和 RNA 聚合酶 II 所利用的启动子位于转录起始位点上游，而 RNA 聚合酶 III 利用的启动子位于转录起始位点下游。

每种启动子都含有相应的一类辅助因子以及供 RNA 聚合酶识别的一套特殊的短保守顺序。RNA 聚合酶 I 和 III 依靠少数的辅助因子识别相对有限的几套启动子。

RNA 聚合酶 II 所利用的启动子中能被转录因子识别的短序列也称作用位点序列，分布在大于 100bp 的区域内，位于转录起始点上游。不同启动子中的短序列及识别它们的因子大多数是共同的，但也有些特别的因子，它们识别特定的一类基因并调节其启动。数目众多的效应因子的作用是促进转录的起始。这些效应因子可以分为三类，前两类与 RNA 聚合酶和启动子的相互作用有关，第三类则在转录调控中起作用。

1）基本因子（basal factor），这类因子在所有启动子起始的 RNA 的合成中都

是必需的，它们和 RNA 聚合酶在转录起始点周围形成复合物，决定起始的位点。

2) 上游因子（upstream factor），是一些能够识别并结合转录起始点上游的一致性短序列的 DNA 结合蛋白。它们普遍存在并且在任何启动子的特定结合位点上起作用，能增强起始的效率，并且使启动子适度地发挥功能。不同的启动子需要不同的一套上游因子，它们对基因的充分表达是必需的。

3) 诱导因子（inducible factor），是在特定的时间或特定的组织被合成或激活的，能控制转录在准确的时间和地点进行的一类因子。这些因子在 DNA 上的结合序列叫做应答元件（response element）。

只含有能被基本因子和上游因子识别的元件的启动子在任何类型的细胞中都能起始转录，这样的基因是细胞中组成型表达的基因（有时也称为持家基因 housekeeping gene），或者是感染广泛宿主并且依赖于宿主转录装置进行增殖的病毒基因。

RNA 聚合酶 II 起始的转录方式多种多样，但它们的共同特征是上游因子和诱导因子结合到转录起始点上游的特定序列上。这些因子与 DNA 结合形成复合物，其中蛋白质与蛋白质的相互作用是很重要的，通过基本转录因子间的相互作用而启动转录。

以上所述的启动子元件序列都位于转录起始点附近。另外还有一些 DNA 序列能增强转录，但这些序列距转录起始点相当远，而且无论是上游或下游，也无论方向如何均能增强转录，这一类元件叫做增强子。增强子元件与启动子元件类似，也由多种多样的标准序列组成，而且这些序列都是相邻的。增强子对 RNA 聚合酶 II 和 RNA 聚合酶 I 催化的转录都有增强作用。

转录起始所必需的，但不是 RNA 聚合酶的成分的任何蛋白质都定义为转录因子。许多转录因子通过识别顺式作用位点起作用，但这不是转录因子作用的惟一方式，一种因子和另一种因子之间以及一种因子与 RNA 聚合酶之间可以识别和相互作用，也可以掺入到起始复合物中。某一种因子只对一种或特定的一套启动子的转录起作用。

真核与原核 mRNA 转录的区别就在于真核转录需要数目巨大的转录因子。真核启动子的定义是含有全部结合位点的区域，该区域支持正常效率的并被适当控制的转录。转录真核 mRNA 的启动子的主要特征在于它是转录因子结合的位点，而 RNA 聚合酶只结合在其周围，并不直接与启动子延伸的上游区域接触。

真核转录调控方式呈正调控，在组织特异性控制下，提供某些转录因子，激活某些启动子。抑制启动子的调控方式很罕见。

真核转录单位通常只包含一个基因，转录终止位点存在编码区下游，但缺少原核中典型的终止区，RNA 聚合酶 I 和聚合酶 III 催化的转录的终止发生在分散的序列上，而 RNA 聚合酶 II 催化的转录终止还不清楚，真核 mRNA 的 3′末端

不是转录终止时形成而是由转录后切割加工产生的。

真核生物中的三种 RNA 聚合酶存在于核内不同的位置，它们的一般性质见表 7-1。

表 7-1　真核细胞中的三种 RNA 聚合酶

酶	位置	产物	相对活性	α-鹅膏素敏感性
RNA 聚合酶 I	细胞核	rRNA	50%～70%	不敏感
RNA 聚合酶 II	核质	hnRNA	20%～40%	敏感
RNA 聚合酶 III	核质	tRNA	～10%	种特异性

转录抑制剂对区分 RNA 聚合酶很有用，不同的抑制剂作用于不同的 RNA 聚合酶，一些抑制物的性质见表 7-2。

表 7-2　转录抑制剂及其作用

抑制剂	作用的酶	抑制剂作用
利福霉素	细菌全酶	结合于 β 亚基阻止起始
链霉素	细菌核心酶	结合于 β 亚基阻止起始
放线菌素 D	真核 RNA 聚合酶 I	结合于 DNA 并阻止延伸
α-鹅膏素	真核 RNA 聚合酶 II	结合于 RNA 聚合酶 II

真核与原核 RNA 聚合酶的主要区别是对 α-鹅膏素的敏感性不同。所有真核 RNA 聚合酶都是大的蛋白质，大于 500 000Da，具有 8～14 个亚基，纯化的酶可以利用 DNA 模板合成 RNA，但不能在启动子处选择起始。

真核 RNA 聚合酶 II 一般由十多个亚基组成（图 7-17）。三个大的亚基与细菌的 RNA 聚合酶有同源性，它们组成基本的催化装置。两个最大的亚基可能携带催化位点，其余三个亚基是三种 RNA 聚合酶所共有的，绝大多数亚基以单拷贝形式存在，有三种亚基以两个拷贝的形式存在。

RNA 聚合酶 II 具有 C 端域（carboxy-terminal domain CTD），由多个重复的一致序列 Tyr-Ser-Pro-Thr-Ser-Pro-Ser 组成，这些序列是 RNA 聚合酶 II 所独有的，这些序列在酵母中重复 26 次而在哺乳动物中重复 50 次左右。重复的次数是重要的，因为缺失典型重复数的一半的突变是致死的。重复序列中的 Ser 和 Thr 是高度磷酸化的，具有非磷酸化亚基的酶称为 RNA 聚合酶 IIa，而具有磷酸化亚基的酶称为 RNA 聚合酶 IIo，CTD 与转录起始反应有关。

通过缺失突变、点突变和足迹法可以鉴定启动子元件。RNA 聚合酶 I 利用的启动子可以分为两部分，－45～＋20 区位于转录起始点周围，称为核心启动子，它足以起始转录；另一序列位于－180～－107 区，称为上游调控元件（up-

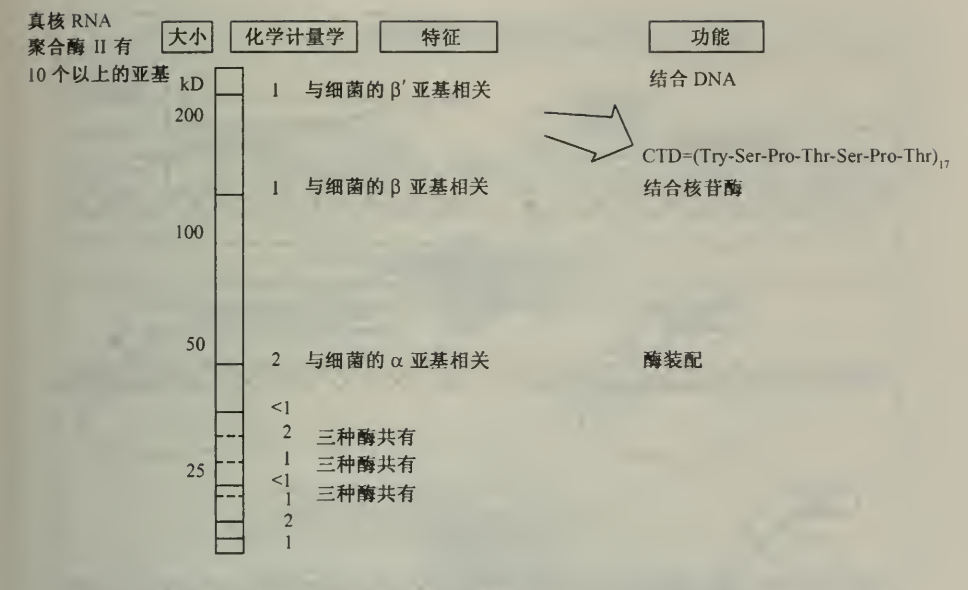


图 7-17　真核 RNA 聚合酶 II 有 10 个以上的亚基

stream control element，UCE），它可以很有效地增强转录效率（图 7-18）。

RNA 聚合酶 I 需要两种辅助因子，UBF1 结合在核心启动子和 UCE 上，SL1 不单独直接与启动子作用，但是一旦 UBF1 与启动子结合，SL1 就可以协同扩展 DNA 覆盖区域。两种因子同时结合以后，RNA 聚合酶 I 就能与启动子结合并起始转录。我们假设在核心启动子上结合的这两种因子与 RNA 聚合酶 I 有直接作用，但我们不知道在 UCE 上结合的这两个同样的因子如何刺激核心区的起始。

UBF1 是单链多肽，它特异地结合于核心启动子和 UCE 中的 GC 丰富区。UBF1 可以识别不同种的模板，如鼠的 UBF1 可识别人的基因的核心启动子和 UCE；SL1 具有种特异性，即人的 SL1 不能识别鼠的基因的核心启动子和 UCE，反过来也一样。

SL1（选择因子）的行为很像细菌中的 δ 因子，它不能特异地与 DNA 结合，但能与 UBF-DNA 复合物结合并使之稳定，从而主管 RNA 聚合酶 I 转录的起始。真核生物 rRNA 转录过程见图 7-19。

RNA 聚合酶 III 利用下游和上游启动子。根据它们被特定的一组转录因子识别的方式不同，这些启动子可分为两类，5S RNA 和 tRNA 的转录启动子是内部启动子、在下游，而 snRNA 转录的启动子在上游（图 7-20）。

tRNA 基因的转录调控区位于转录单元内的转录起始点之后。tRNA 基因中有两个高度保守的启动子序列，即 A 盒（A box，5′-TGGCNNAGTGG-3′）和 B

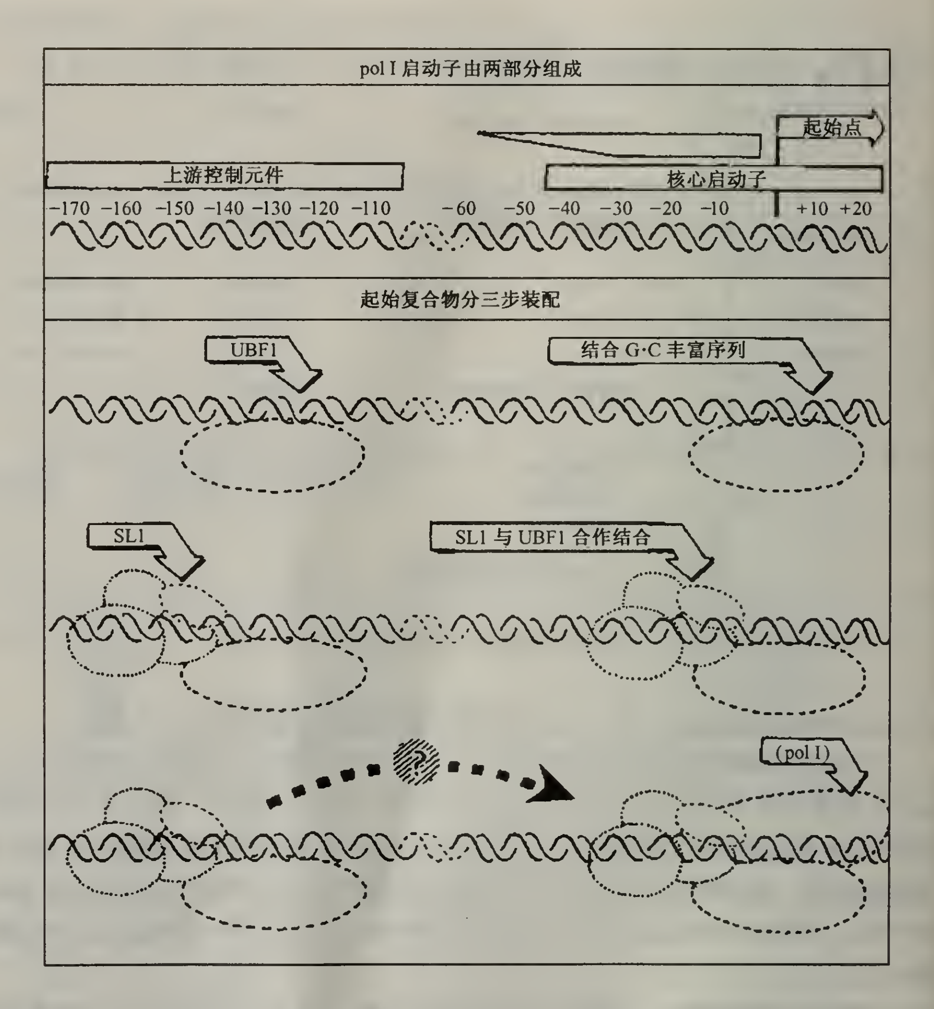


图 7-18 RNA 聚合酶 I 的启动子由一个核心启动子组成

距上游控制元件 70bp，UBF1 可与两个区域结合，之后 RNA 聚合酶 I 与核心启动子结合。

盒（B box，5′-GGTTCGANNCC-3′）。这两个序列同时也编码 tRNA 自身的重要序列，即编码 tRNA 的 D 环和 TΨC 环。这意味着 tRNA 内的高度保守序列同时也是高度保守的启动子序列。

RNA 聚合酶 III 起始转录所必需的两个复杂的转录因子 TFIIIB 和 TFIIIC 已鉴定出来了（图 7-21）。TFIIIC 可以与 tRNA 启动子的 A 盒和 B 盒结合，TFIIIB 则可以和 A 盒 50bp 上游序列结合。TFIIIB 由三个亚基组成，其中一个是 TBP（TATA box binding protein），三种 RNA 聚合酶都需要；第二个亚基被称为 BRF（TFIIIB 相关因子，它与 RNA 聚合酶 II 起始因子 TFIIB 同源）；第三个亚基是 B″。TFIIIB 没有序列特异性，因为它的结合位点似乎是由 TFIIIC 与

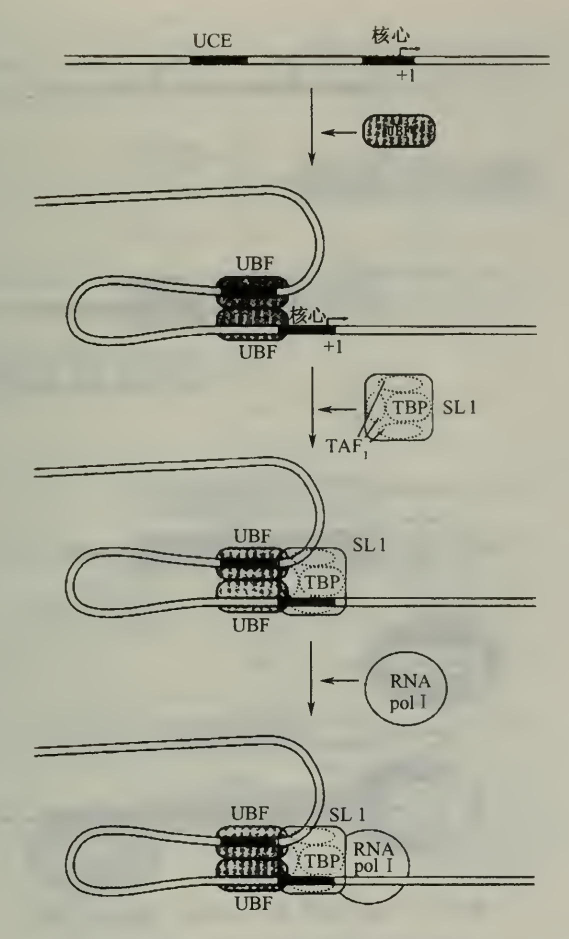


图 7-19　rRNA 转录起始的图解模型

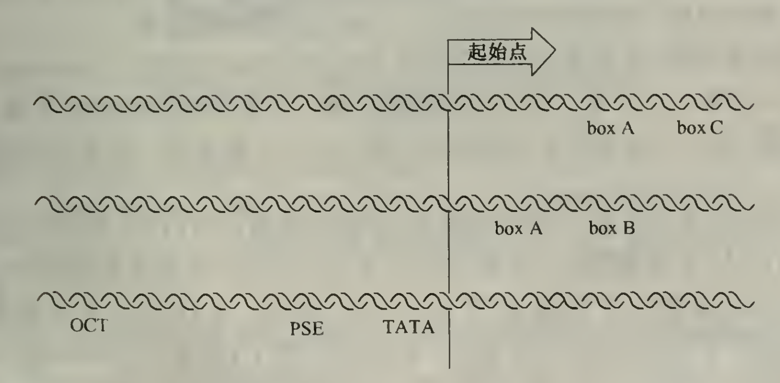


图 7-20　RNA 聚合酶 III 利用的启动子可能由两部分组成

起始点下游由分开的 boxA 和 boxC 或 boxA boxB 组成，

上游可能由分开的顺序（OCT PSE TATA）组成

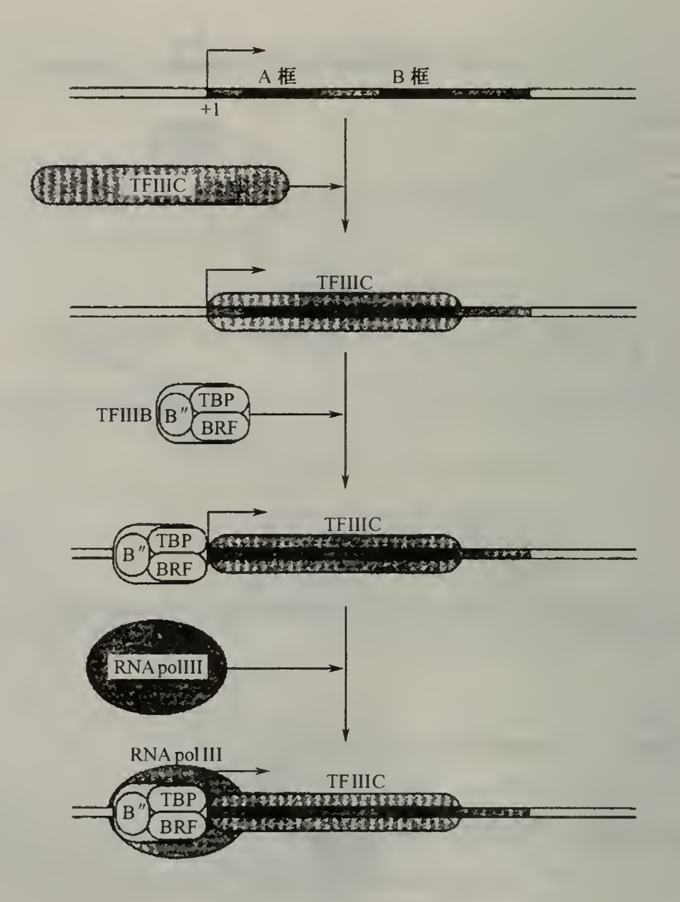


图 7-21　真核生物 tRNA 启动子转录的起始

DNA 的结合位点所决定的。TFIIIB 可以促进 RNA 聚合酶 III 结合并起始转录。TFIIIB 结合以后，TFIIIC 就相当于一个为其定位的装配因子。

RNA 聚合酶 III 利用的另一类启动子是 5S rRNA 启动子。5S rRNA 基因的启动子上含有一个 A 盒，位于内部调控区域，它处于转录起始点下游 + 50 至 + 65之间。另一个 C 盒位于转录起点下游 + 80～ + 99 之间，这两个序列都非常重要。

5S rRNA 启动子的 C 盒是特异的 DNA 结合蛋白 TFIIIA 的结合位点（图 7-22）。TFIIIA 作为装配因子使得 TFIIIC 可以与 5S rRNA 启动子相互作用。A 盒可能也稳定 TFIIIC 与 DNA 的结合，一旦 TFIIIC 与 DNA 结合形成复合物，TFIIIB 就能够与该复合物起作用并使 RNA 聚合酶 III 结合进来而起始转录。

RNA 聚合酶 II 自身不能启动转录，它必须依赖于辅助的转录因子，聚合酶与这些转录因子组成一个基本转录装置，这个装置对于任何启动子都是必需的。RNA 聚合酶 II 结合的启动子含有多种短的顺式作用元件，每一种元件被一种反

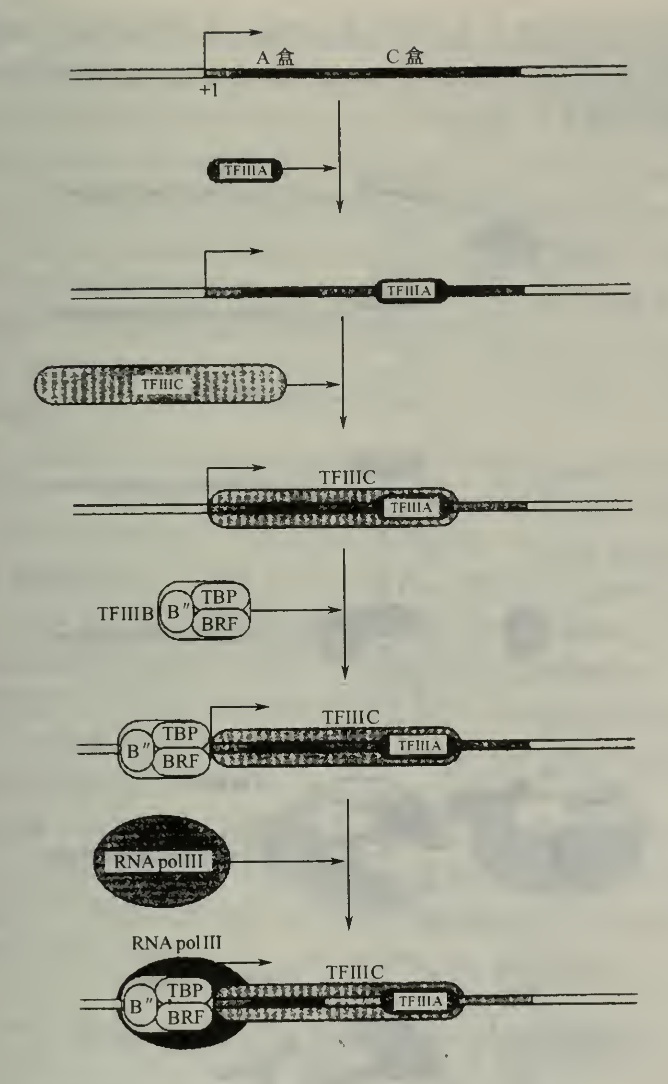


图 7-22　真核生物 5S rRNA 启动子转录的起始

式作用因子所识别。这些顺式作用元件位于 TATA 盒的上游，并且可能以两种方向和不同的距离存在。靠近起始位点的 TATA 盒和直接位于起始位点上的起始子区（initiator region）负责起始位点的准确选择，远上游的元件决定启动子的效率。

启动子可被增强子激活而增强转录，增强子距离它所作用的启动子很远，它可以从两个方向在一个基因的两边起作用。增强子由一套元件组成，这些元件组织（orgnization）得比较集中。有些元件为增强子和启动子所共有。增强子通过与启动子上装配的蛋白质复合物相互作用而发挥功能。

已被鉴定 RNA 聚合酶 II 起始转录所必需的一系列转录因子、纯化和克隆。这些因子被命名为 TFIIA、TFIIB（transcription factor IIA，IIB）等。它们以特定的次序组装到启动子上（图 7-23）。

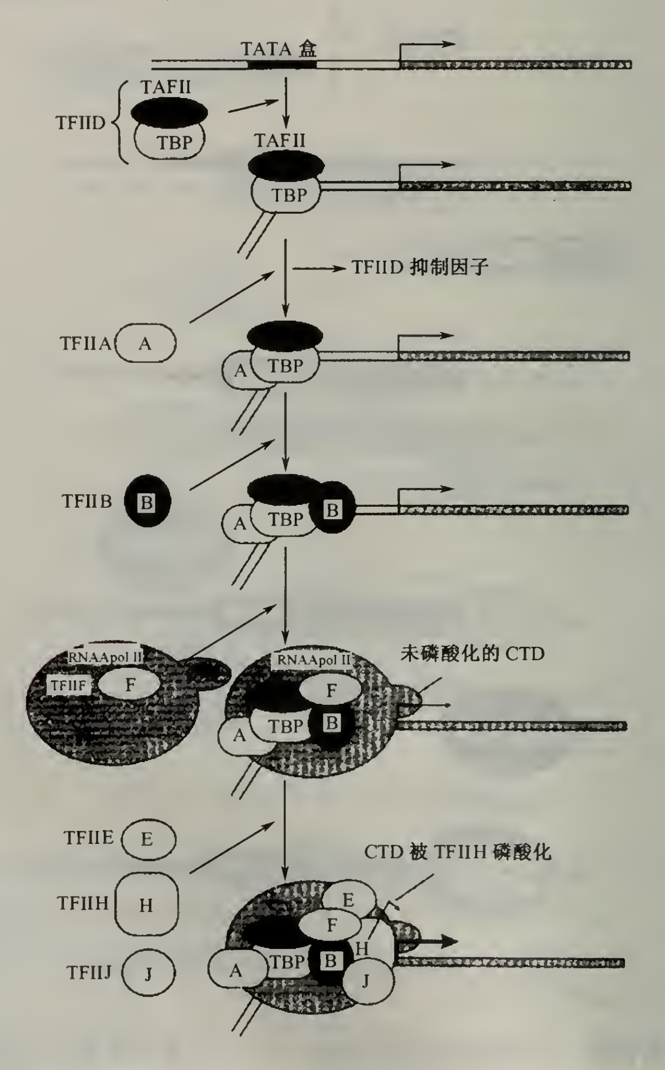


图 7-23　RNA 聚合酶 II 转录起始复合体在一含有 TATA 盒的启动子上的组装

RNA 聚合酶 II 起始复合物组装的第一步是 TFIID 识别相应序列并与之结合。TFIID 是由 TBP 和 TAF II（TBP 相关因子）组成的复合物。TBP 与 TATA 结合。研究表明在哺乳动物细胞中，TBP 与 TATA 结合后，接着至少有 8 个 TAFII 加入进来形成 TFIID。

TBP 在真核生物的三种转录复合物（SL1、TFIIIB 和 TFIID）中都存在，

在转录起始中起重要作用。分析过的所有真核生物 TBP 均是单体蛋白并都有一个由 180 个残基构成的非常保守的 C 端结构域，这个保守的结构域在体内转录中所起的作用与全长蛋白质所起的作用相同。TBP 是一个完全二元对称结构，作用于 DNA 的小沟。TBP 内部结构与 TATA 结合，外部结构与其他蛋白质因子相互作用。TATA 结合区发生突变将抑制 RNA 聚合酶 II 的转录起始。

TFIIA 与 TFIID 的结合能增强 TFIID 与 TATA 盒的结合，并稳定 TFIID-DNA 复合物。体外转录研究发现，TFIID 得到纯化后，就不再需要 TFIIA。所以在完整细胞中，TFIIA 的作用似乎是抵消抑制因子诸如与 TFIID 相关 DR1 和 DR2 的作用，很可能 TFIIA 与 TFIID 的结合阻碍了这些抑制因子的结合，让组装过程得以继续。

TFIID 一旦结合到 DNA 上，TFIIB 就会与 TFIID 结合，而且 TFIIB 又可与 RNA 聚合酶结合。TFIIB 起一个桥梁作用，它允许 RNA 聚合酶以及携带聚合酶的 TFIIF 加入到复合物中来。TFIIF 含有两个亚基，大亚基 RAP74 具有依赖 ATP 的 DNA 解旋酶活性，小亚基 RAP 38 和 RNA 聚合酶 II 紧密结合。复合物装配完成后，RNA 聚合酶 II 要从复合物中释放出来才能起始转录。负责把 RNA 聚合酶 II 释放出来的是 TFIIH，它的一个亚基具有蛋白质激酶活性，能使 RNA 聚合酶 C 端重复序列（CTD）YSPTSPS 中的 S 和 T（丝氨酸和苏氨酸）磷酸化，磷酸化使 RNA 聚合酶 II 发生变构并与 TFIIF 分离而从复合物中释放出来，沿着 DNA 模板移动开始转录。

RNA 聚合酶 II 的转录终止能力目前还没有完全研究清楚，转录产物的 3′末端是切割产生的。位于切割位点上游 10～30 个碱基的 AAUAAA 序列是切割和加多聚 A 的信号。转录基本复合物中的一种内切核酸酶和 poly（A）聚合酶及其他因子（这些因子具有识别 AAUAAA 信号的专一性）相结合。

二、真核 mRNA 分子 5′端和 3′端的结构

1. 真核 mRNA 分子 5′端“帽子”

真核 mRNA 分子 5′端是由 7-甲基鸟嘌呤核苷酸与 2′-O-甲基鸟嘌呤核苷酸或 2′-O-甲基腺嘌呤核苷酸以 5′-5′焦磷酸连接而封闭的。5′端“帽子”的结构单位是（7-MeG）-5′-ppp-5′-［2′-OMe（G/A)］-3′-p-5′-核苷酸-3′-p。

图 7-24 是真核 mRNA 分子的 5′端结构，有三种结构类型的帽子（m^7G5' pppNp，m^7G5'ppp5′Nmp-Np 和 m^7G5'ppp5′NmpNmp），在帽子 0 中，核糖不被甲基化；在帽子 1 中有一个核糖被甲基化；在帽子 2 中有两个核糖被甲基化。除真核生物 mRNA 外，高等动植物病毒也有 5′帽子结构，有的病毒具有特殊的帽子结构，这些帽子结构都是转录后加上去的。

帽子1和2甲基化

帽子2处甲基化

图 7-24　真核 mRNA 的 5′端帽子结构

帽子结构对于 mRNA 的翻译很重要，它可使 mRNA 免遭外切核酸酶的降解，同时它对于 mRNA 与核糖体的结合以及翻译起始区的辨认也很重要。呼肠孤病毒（reovirus）mRNA 和水泡性口炎病毒（vesicular stomatitis virus VSV）mRNA 都有甲基化的 5′末端帽子结构。带甲基化 5′端的 mRNA 比 5′端不甲基化的 mRNA 在麦胚系统中合成蛋白质的速度快得多。如果用化学方法除去 m^7G，病毒 RNA 和兔珠蛋白 mRNA 就失去了绝大部分模板活性。

5′端帽子结构与蛋白质翻译的起始有关。无 m^7G 的呼肠孤病毒不能与 40S 亚基形成复合物。m^7G 可能是蛋白质合成起始的辨认信号的一部分，如小鸡卵白蛋白 mRNA，其 5′末端的帽子结构附近区域能形成发夹结构，这样的结构使 5′端帽子结构与蛋白质合成的起始密码子在空间上靠近了，这种 5′端附近的发夹结构在其他真核 mRNA 中也存在，因此帽子结构与蛋白质合成的启动有关。

2. 真核 mRNA 分子 3′末端尾巴

绝大多数（不是全部）的真核 mRNA3′端有 poly（A）尾巴，它是 mRNA 转录后早期由腺苷酸聚合酶加上去的，其长度一般是 20～200 核苷酸，poly（A）不能加到初级转录产物上，因为在正常情况下转录是通过 poly（A）信号的，这样在加上 poly（A）之前，内切核酸酶必须发生作用。据目前了解的情况来看，poly（A）尾巴可能的功能有：①对稳定 mRNA 有一定作用。实验证明，去除 poly（A）以后珠蛋白 mRNA 的模板活性很快消失，这说明 poly（A）的存在与 mRNA 的抗降解有关，但聚腺苷酸如何影响 mRNA 的降解目前还不清楚。②poly（A）还与 mRNA 顺利通过核模进入细胞质的过程有关，有人认为 poly（A）与病毒的侵染性也有关。③poly（A）与翻译调控有关。有些 mRNA 没有 poly（A）尾巴，如组蛋白 mRNA，这种 mRNA 的 3′端有特殊的二级结构，3′末端的形成依赖于此二级结构。RNA 的合成在茎环结构处终止，妨碍双股的茎形成的突变阻止 RNA3′端的形成，修复突变可以恢复二级结构并且也恢复形成

3′端的能力。因此推断，3′端的二级结构比准确的序列对于 3′端形成更为重要。形成茎环二级结构的链是 DNA 上的编码链而不是模板链。

体外实验证明，3′端的正确形成需要三种活性成分的作用，一种是热不稳定因子（功能尚不清楚）；第二种是与 3′切割位点上游发卡结构结合的因子；第三种是一个 56 个核苷酸的 RNA 称为 U7 RNA，它与 3′端切割位点下游的一段短序列互补。U7 snRNA 与组蛋白 H3 mRNA 的相互作用如图 7-25。在几种不同物种来源的 H3 mRNA 中，与 U7 snRNA 互补的上游发夹结构和下游顺序是保守的。U7 snRNA 的序列和结构从海胆到人都是很保守的。

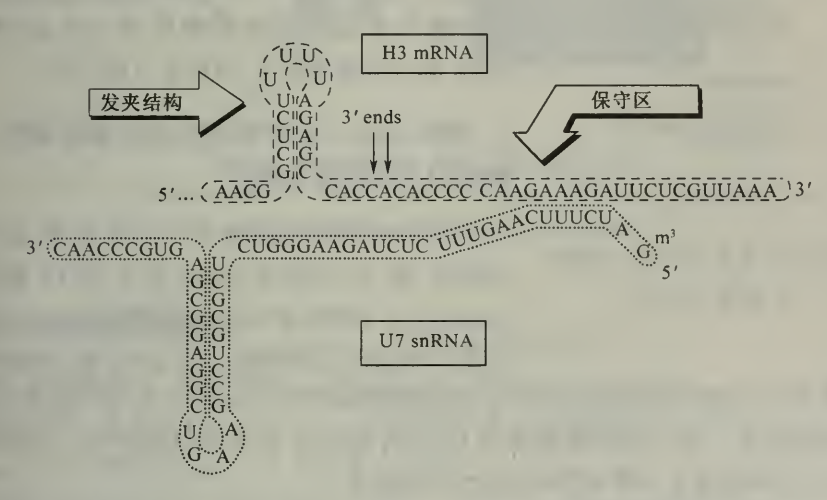


图 7-25　组蛋白 H3 mRNA 3′端的产生依赖于两个特性

3′端有一发夹结构和一个与 U7 snRNA 碱基配对的顺序。

组蛋白 mRNA 的下游一致性序列的突变降低了它与 U7 snRNA 配对的能力，并且抑制 3′端的加工，U7 snRNA 相应突变可以恢复配对能力，也能恢复 3′端的加工。因而推断 U7 snRNA 功能的发挥要通过碱基配对，但还不知道具体步骤，可能是终止也可能是切割。

三、真核生物 mRNA 的加工

真核生物 mRNA 的初级转录产物必须经历几个加工步骤才能成为成熟的 mRNA，包括三个过程：5′端加帽、3′端切割及多腺苷酸化和 RNA 拼接。

1. 5′端加帽

新生 RNA 起始合成大约 25～30 个碱基后，7-甲基鸟嘌呤就被加到 5′末端。RNA 加工过程中，这个步骤的起始由二聚体加帽酶催化，该酶与 RNA 聚合酶 II

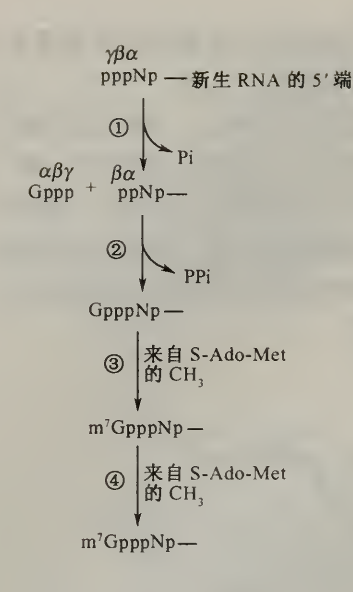


图 7-26 新生 RNA 的 5′端加 7′-甲基 G（m^7G）

的 CTD 的磷酸化有关。CTD 在转录起始期间被磷酸化，加帽酶与 RNA 聚合酶 I 和 II 并无相互作用，它专门针对 RNA 聚合酶 II 转录产生的产物进行催化反应。

加帽酶的一个亚基催化 RNA 聚合酶 II 合成的新生 RNA 的 5′末端除去 γ 磷，另一亚基将来自于 GTP 的 GMP 转移到新生 RNA 的 5′-二磷酸上，产生鸟嘌呤 5′-5′-三磷酸结构，最后由单独的酶将甲基转移到鸟嘌呤的 N-7 位和新生 RNA 的 5′端核糖的 2′-O 位上（图 7-26）。

2. 前体 RNA 与含有保守的 RNA 结合域的 hnRNA 结合蛋白相联系

蛋白质编码基因转录产生的新生 RNA 和 mRNA 加工过程中的中间产物体系统称 pre-mRNA（前体 RNA）。真核细胞核内的 RNA 不是以自由形式存在的。从 RNA 聚合酶起始转录产生 RNA 起直到成熟的 RNA 被运输到细胞质中，RNA 一直与一套丰富的核蛋白质联系在一起，这些蛋白是与 hnRNA 结合的主要蛋白质组分，hnRNA 是 pre-mRNA和其他大小不等的核内 RNA 的统称。

不同的 hnRNA 结合蛋白与新生 pre-mRNA 分子的不同区域结合，如 hnRNA 蛋白质 A1，C 和 D 优先结合在内含子 3′端的嘧啶丰富区。

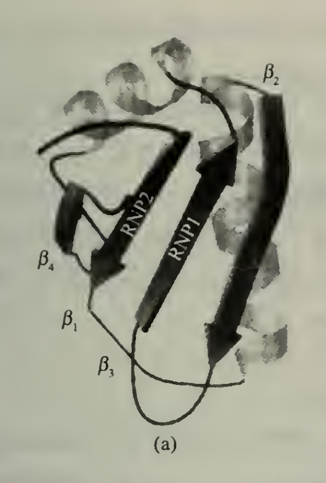
与转录因子类似，几乎所有的 hnRNA 结合蛋白都有特定的分子结构，它们含有一个或多个 RNA 结合域，而且至少有一个结构域用于与其他蛋白质相互作用。

RNA 结合蛋白含有与 DNA 结合蛋白相同的锌指结构域。

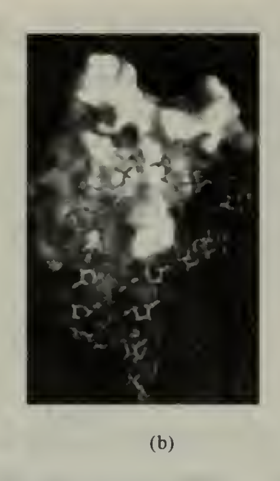
RNA 结合蛋白结构域也称为 RNA 结合域（RBD），它在 hnRNA 结合蛋白中普遍存在，一个大约 80 个残基的结构域存在于任一 RNA 结合蛋白中。RNA 结合蛋白含两个高度保守的区域（RNP1 和 RNP2）。

该 RNA 结合蛋白结构域由四个 β 折叠和两个 α 螺旋组成。RNP1 和 RNP2 在两个中心 β 链上并列，而且其侧链与 RNA 单链区有多处结合。单链 RNA 环穿越 β 折叠表面并恰好进入一个存在于 $β_2$、$β_3$ 链结合蛋白质环与 RNA 结合蛋白 C-末端区域之间形成的横沟中（图 7-27）。

RGG box 是 hnRNA 结合蛋白中发现的另一个 RNA 结合结构域，它含有五个 Arg-Gly-Gly（RGG）重复，这种富含精氨酸的特性与 λ 噬菌体 N 蛋白和 HIV Tat 蛋白的 RNA 结合域相似。



(a)



(b)

图 7-27 UIA 蛋白和 RNA 复合物的结构

(a) RNP 结构域简图，保守的 RNP1 和 RNP2 位于两个中等 β-链中。

(b) X 射线结晶学测定的 UIA 蛋白质 RNA 复合物。

3．Pre-mRNA3′端的加工

在动物细胞中，除组蛋白 mRNA 以外，所有 mRNA 都有 3′-poly（A）尾巴。早期脉冲标记腺病毒和 SV40 RNA 的研究表明，病毒初级转录产物可以延伸到加 ploy（A）信号的下游，这个结果说明附加 ploy（A）的 3′-OH 末端是由核酸内切酶切割产生的，但是预期的下游 RNA 片段由于在体内降解得太快以致无法检测到。然而，用 Hela 细胞核抽提物进行的体外系统加工反应中检测到了预期的切割产物，从而进一步验证了这种内切核酸酶切割机制。

动物细胞 cDNA 克隆的早期测序结果显示在所有含 AAUAAA 序列的 mRNA 的 ploy（A）尾巴上游 10～35 个核苷酸处，如果编码 AAUAAA 的模板变成其他任何序列时，该基因的转录产物就失去添加 ploy（A）的能力，由该突变模板转录的未加工产物很快就被降解了。在动物细胞中，切割位点下游的第二个信号对于多数前体 mRNA 的切割和 poly（A）化是必需的。该信号是 GU 丰富区或 U 丰富区，在切割位点周围的 50 个核苷酸之内。

前体 mRNA 的切割和附加多聚腺苷酸所需的蛋白质逐渐被纯化和鉴定，于是科学家提出了图 7-28 中的模型。根据这个模型，一个由四条链组成的 360kDa 的切割和多腺苷酸化特异因子（cleavage and polydenylation specificity factor CPSF）首先与第一个 poly（A）信号上游 AU 丰富区形成不稳定复合物，然后，三个外加蛋白与 CPSF-RNA 复合物结合，这三个外加蛋白包含一个称为切割刺激因子（cleavage stimulatory factor CStF）的 200kDa 异源三聚体，一个称为切割因子（cleavage factor I CFI）的 150kU 异源三聚体和另外一个切割因子（CF

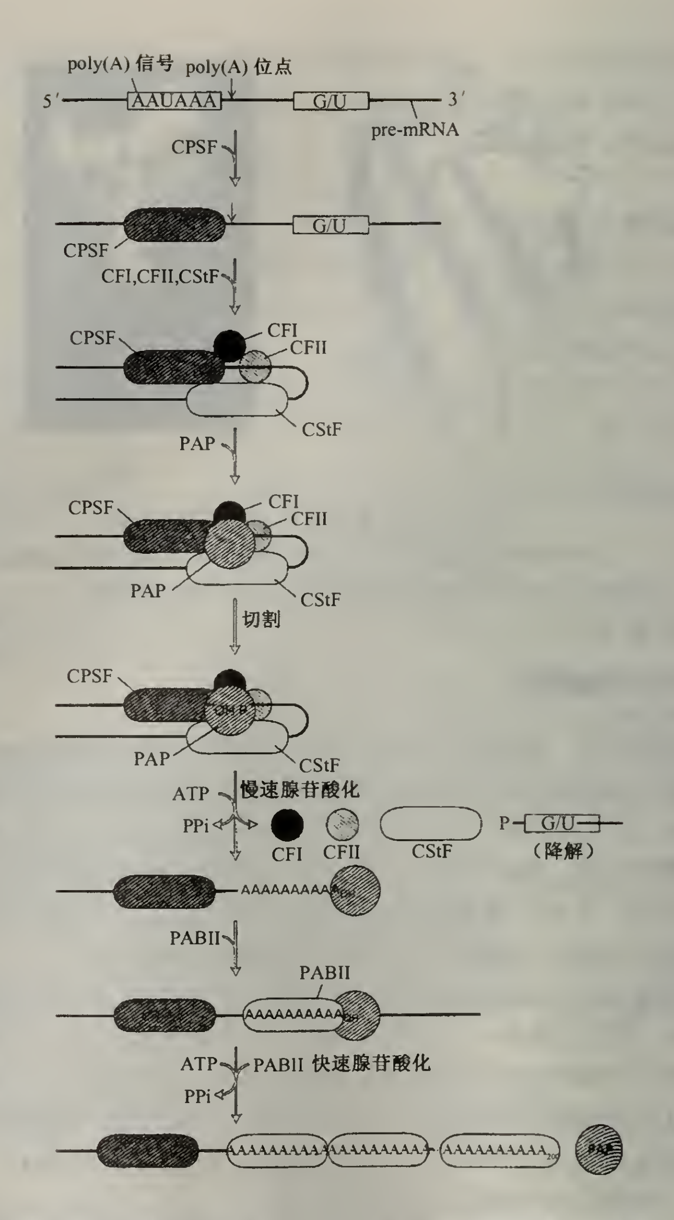


图 7-28　哺乳动物细胞中的 Pre-mRNA 的切割和多腺苷酸化的模型

II)。CStF 和 poly（A）信号下游 GU 丰富区或 U 丰富区的相互作用能稳定该多蛋白复合物。最后，poly（A）聚合酶［poly（A）polymerase PAP］结合到该复合物上，并切割产生自由 3′ 末端后迅速形成 poly（A）。

真核生物前体 mRNA3′端的切割和 poly（A）化过程中的多蛋白装配与 DNA 分子 AT 丰富区的 TATA box 上转录起始复合物的装配过程类似。多蛋白复合物

的装配过程中，贯穿着特异性的蛋白质-核酸、蛋白质-蛋白质相互作用网络的协同作用。

加 poly（A）紧接在 poly（A）位点切割之后，加 poly（A）分两个时期，第一时期是慢速加上 12 个 A，接着快速加上 200～250 个以上 A。快速期需要含 RNA 结合蛋白结构域的 poly（A）结合蛋白的多拷贝结合，该蛋白定名为 PAB II 以区别于结合在细胞质 mRNA poly（A）尾巴的 poly（A）结合蛋白。PAB II 与 PAP 起始加上的短 poly（A）尾结合，可以进一步促进 PAP 催化的 poly（A）化反应（图 7-28）。

4．RNA 拼接

所有真核 mRNA 分子初级转录产物中都有插入序列即内含子（intron），内含子把编码序列隔开，真核 mRNA 分子必须经过加工去除内含子并拼接后才能作为蛋白质翻译的模板。RNA 的拼接（RNA splicing）发生在细胞核内。5′端帽子和 3′端尾巴形成之后，内含子一个一个被切除，编码序列（称为外显子 exon）相互连接形成成熟的 mRNA 这个过程就叫 RNA 的拼接。拼接后的 mRNA 仍有 5′端帽子和 3′端尾巴。

采用异源双链技术，在电镜下可以清楚地看到不参与形成成熟 mRNA 分子的双链 RNA 部分（intron）（图 7-29）。

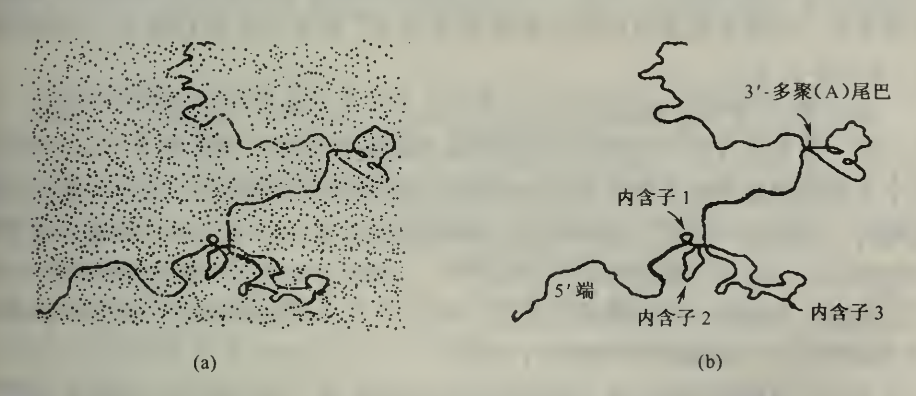


图 7-29　成熟 mRNA 与 DNA 杂交后的电镜照片

RNA 的拼接具备以下特点：切割必须在特定位点上进行，即切割时必须识别出内含子和外显子连接处的特殊碱基序列，否则即使错切一个碱基，阅读框也会发生改变。

分析许多内含子的碱基序列发现，每个内含子的 5′端都是 GU，3′端都是 AG（图 7-30）。

mRNA 加帽和加 poly（A）之后，在核内发生拼接反应。由于初级转录产物

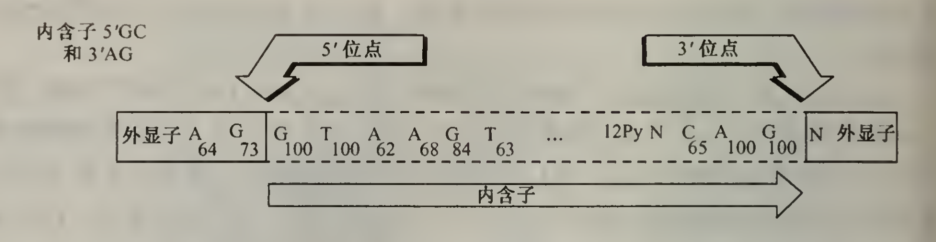


图 7-30　内含子两端的序列

被逐步加工，所以核内 RNA 分子的大小很不一致。故称为异质核内 RNA (hnRNA),加工完成以后的成熟 mRNA 才能从核内转运到细胞质中，然后被翻译。是什么因素阻止初级转录产物和部分加工产物被转运到细胞质目前还不清楚。一般认为拼接在核膜上发生，接着成熟 mRNA 通过核孔转运到细胞质内。

所有内含子都具有相同的末端，那么拼接究竟如何发生的呢？内含子的切除是有明显次序的还是随机的？如果按专一的次序进行，那么将仅仅产生单一的中间体；如果是随机进行的，那么将存在各种中间体。在对少数基因进行的研究中观察到内含子不是以一个单一次序切除的，而是以选择次序切除的。例如：鸡卵蛋白基因有 7 个内含子（A～G），如果随机切除应产生 300 种中间体，但实际上只见到少数几种中间体。失去一个内含子的中间体常常是失去了 E 或 F，接着两个都失去，即首先失去的内含子常常是 E 或 F，接下去失去 G 或 D，然后再是 B，最后是 A 或 C。

核内 RNA 拼接所需要的只是在 5′位点、3′位点和分支位点的三个短的一致性序列。分支位点的一致性序列在酵母中是高度保守的，该序列为 UACUAAC。这个序列在高等真核生物中是不太保守的，但是具有可识别的一致性，即各位置的嘌呤、嘧啶是一致的，该序列为：Py80NPy80Py87Pu75APy95。绳套拼接过程分三步进行：①在 5′位点切开，内含子 5′-G 和在分支位点一致性序列中的 A 以 5′-2′磷酸二酯键结合形成绳套。②在 3′端切开并且连结外显子，内含子以绳套形式被释放。③绳套解开成线状内含子。

分支位点的靶序列位于 3′拼接位点上游 18～40 个核苷酸处，分支位点靶序列是用来鉴别将与 5′拼接位点连结的最靠近 3′拼接位点的靶点。在酵母中，如果 UACUAAC 突变，拼接反应的任何一步均不能发生。5′拼接位点突变后不影响绳套的形成。如果内含子开头的 G 变为 A，那么切割在左边的外显子进行，绳套可以形成，但到此为止。如果 5′-2′分支序列不正确，则在它的 3′端下游不发生切割，其他的一些突变也容易阻断该切割反应。

根据 3′位点的作用，内含子分两类，一类内含子的 3′位点突变不影响 5′位点的加工和分支位点。另一类内含子的 3′位点的变异会妨碍 5′拼接位点的切割。

拼接的第一步是通过 5′拼接位点上 UACUAAC 中的 A 的亲水攻击，第二步

是由第一步释放的 3′端再去攻击 3′拼接位点，在这个过程中，磷酸二酯键的数目是保守的，在外显子和内含子（exon 和 intron）拼接位点原来有两个磷酸二酯键，在拼接过程中一个被两个外显子之间的 5′-3′磷酸二酯键取代，另一个被形成绳套中的 5′-2′磷酸二酯键所代替（图 7-31）。

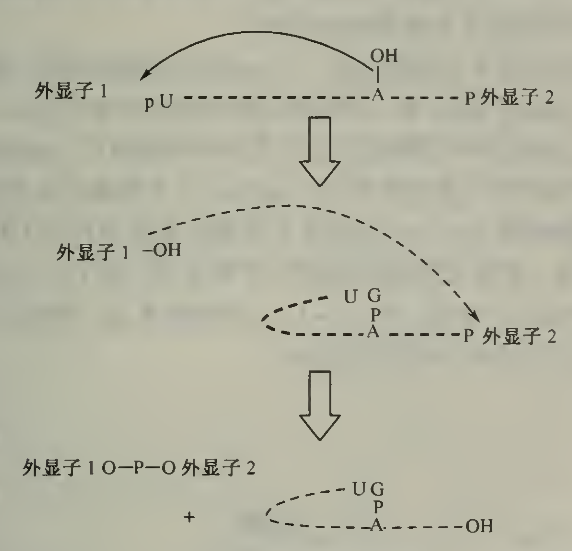


图 7-31　在一个自由的 3′-OH 末端通过两次转酯反应，攻击一个磷酸双酯键，发生核剪接

真核生物细胞核和细胞质中含有大量的小分子 RNA。在高等生物中，小分子 RNA 大小大约 100～300 个核苷酸，在酵母中可长达 1000 个核苷酸左右，它们在细胞中相当丰富达每细胞 105～106 分子，种类也很多，但它们的浓度太低而不能直接检测，不过它们与加工过程中有关的功能已被鉴定。核内的小分子 RNA 叫 snRNA（small nuclear RNA），细胞质中的小分子 RNA 叫 scRNA（small cytoplasmic RNA），在天然状态下，它们以核糖核蛋白粒子形式（snRNP 和 scRNP，有时候也称为 snurp 和 scyrp）存在。一个 snRNP 通常只含有单个 RNA 分子和 10 种左右的蛋白质，有些蛋白质是某一特定 snRNP 所特有的，而有些蛋白质是几种类型的 snRNP 所共用的。一些公共蛋白质被自身免疫抗血清所识别，称作 anti-sm；这些蛋白质一定与自身免疫识别有关系，但它们与自身免疫病的关系目前还不清楚。

通过突变研究发现一些较丰富的 snRNA 在酵母的生活中不是必不可少的，而与拼接有关的 snRNA 是必需的，它们的基因突变是致死的。

一些 snRNA 与拼接有关，某些 snRNA 具有与 5′和 3′拼接位点以及分支靶序列互补的序列，它们之间可以进行碱基配对，并在拼接中起重要作用。

拼接装置包括几种 snRNP，并有大量附加的单个蛋白质，常称作拼接因子，

这些 snRNP 包括 U1、U2、U5 和 U4/U6，它们是根据细胞内存在的 snRNA 的名称而命名的，U1、U2 和 U5 snRNP 各含单个 RNA 分子和几种蛋白质，U4/U6 snRNP 含有 U4 和 U6 两条 RNA 分子。各种 snRNP 中蛋白质的总和大约有 40 种，其中的几种可能直接与拼接有关，而有些仅仅是维持结构所需要的，或者是 snRNP 粒子的相互作用或装配所需。

人的 U1 snRNP 含有 8 种蛋白质，U1 snRNP 可能形成的二级结构含几个域如图 7-32 所示，在 U1 RNA 中，茎环结构提供蛋白质结合位点。在体外，完整的 U1 snRNP 可以结合到 5′拼接位点上并且只有完整的 U1 snRNP 颗粒才具有结合能力。用突变分析可以直接检查 U1 snRNA 与 5′拼接位点配对的必要性。图 7-33 中，野生型腺病毒 12S mRNA前体 5′拼接位点的序列与 U1 RNA 的六个位置配对后进行拼接。如果 12S mRNA前体 5′端突变，即 GG→AU，则拼接不能进行。但当 U1 RNA 也突变，即 C→U，则拼接恢复，因此 U1 RNA 与 12S mRNA 5′拼接位点互补对于拼接是必需的。

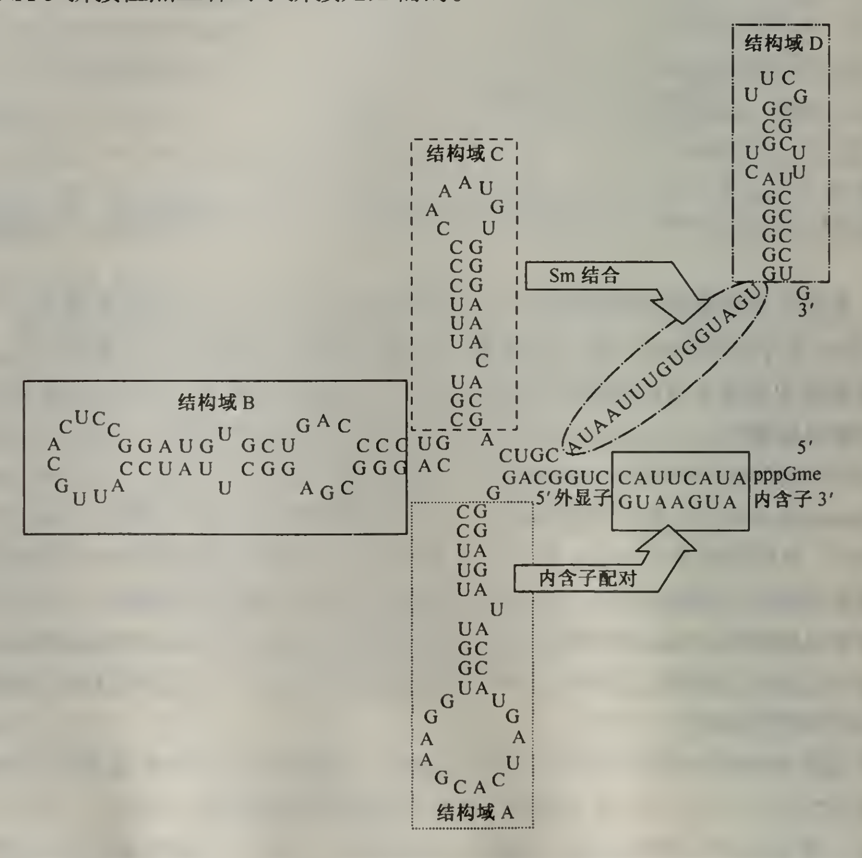
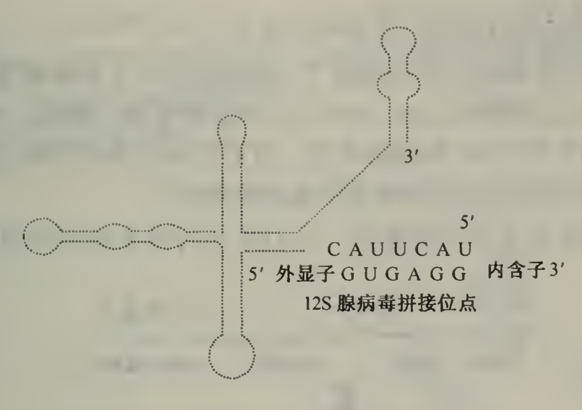
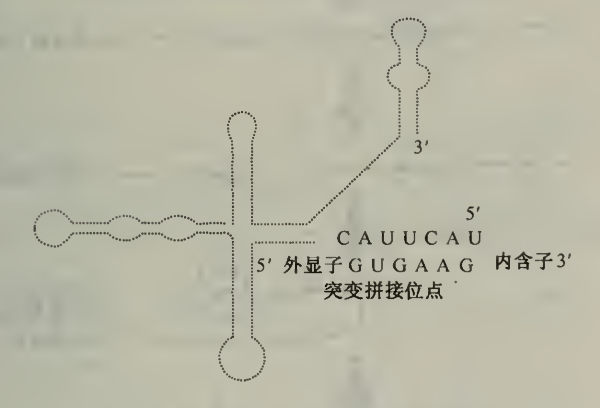


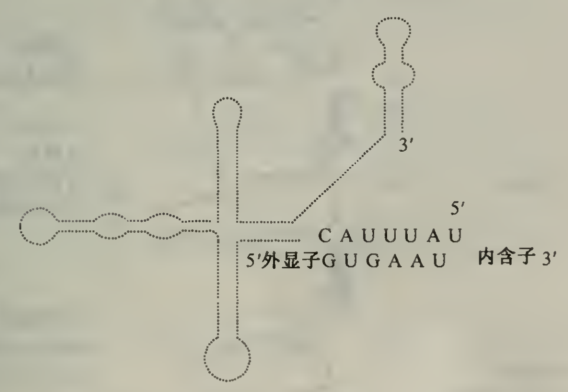
图 7-32　U1 RNA 有一些碱基配对的结构形成一些结构域，其 5′端单链可与 5′拼接位点配对



野生型 U1 RNA 和 12S mRNA 前体正常拼接



野生型 U1 RNA 和 12S mRNA 前体不拼接



突变 U1 snRNA 和突变的 12S mRNA 拼接恢复

图 7-33 消除 5′拼接位点功能的突变可被 U1 snRNA 中与其碱基配对处的突变所抑制

U2 RNA 含有与分支位点互补的序列，用抗 U2snRNA 的抗体进行免疫沉淀可直接鉴定出 U2 snRNP 和包括分支位点在内的内含子形成的复合物。snRNA

近 5′端的序列与内含子中的分支序列互补。

与拼接有关的 snRNP 和拼接因子一起形成一个大的颗粒复合物，称为拼接体（spliceosome），拼接体为 50～60S 大小的核糖核蛋白颗粒。有些 RNA 拼接蛋白与依赖于 ATP 的 RNA 解链酶有关，它们在拼接体中 RNA-RNA、RNA-蛋白质相互作用或在结构的创立和释放中起重要作用。

几种 snRNA 的序列已经测定，它们的共同特点是可以形成不同的结构域

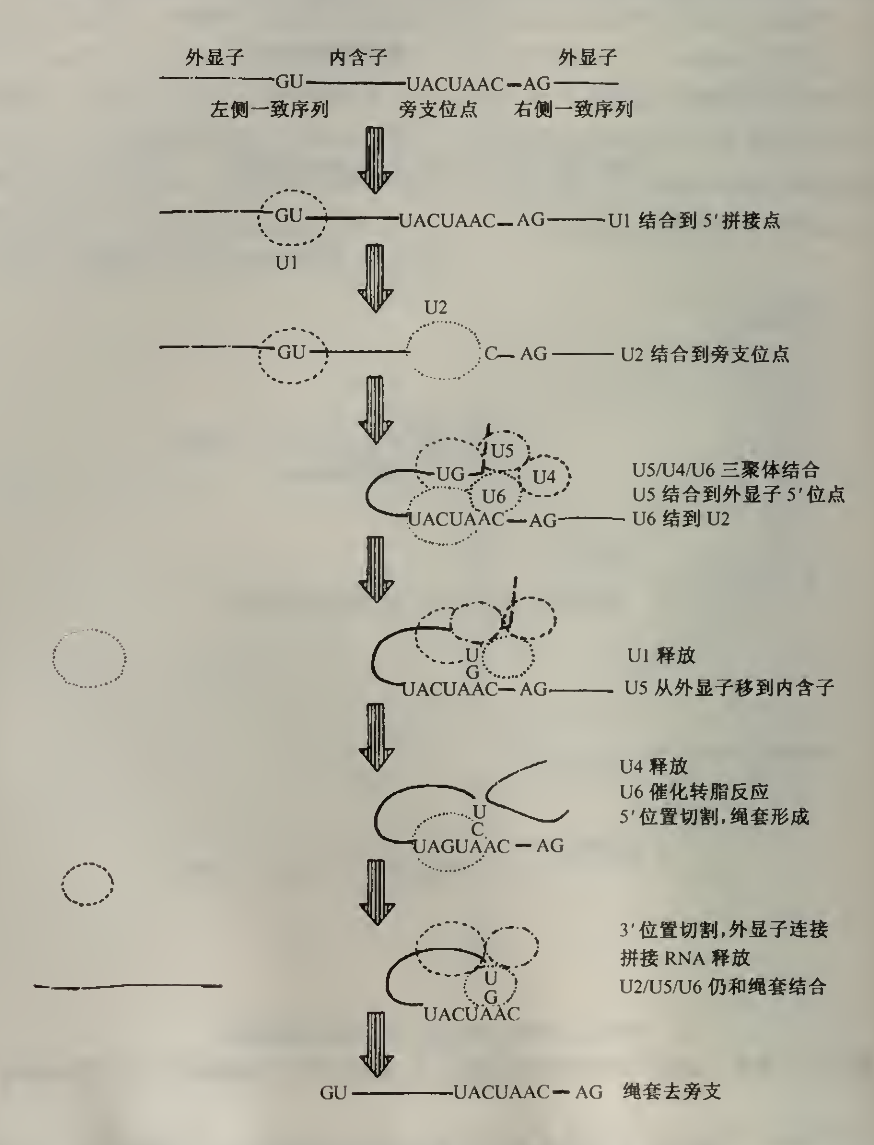


图 7-34　拼接通过不连续步骤进行，该过程中形成拼接体，包括识别一致性序列的各成分的相互作用

(domain)（图 7-32），而且都有与前体 mRNA 5′端位点、3′端位点、分支靶序列以及不同的 snRNA 之间互补的碱基序列。这些特点使它们在拼接中起作用。拼接过程见图 7-34，下面是几种拼接模式：

1）I 组内含子自我拼接（图 7-35）；

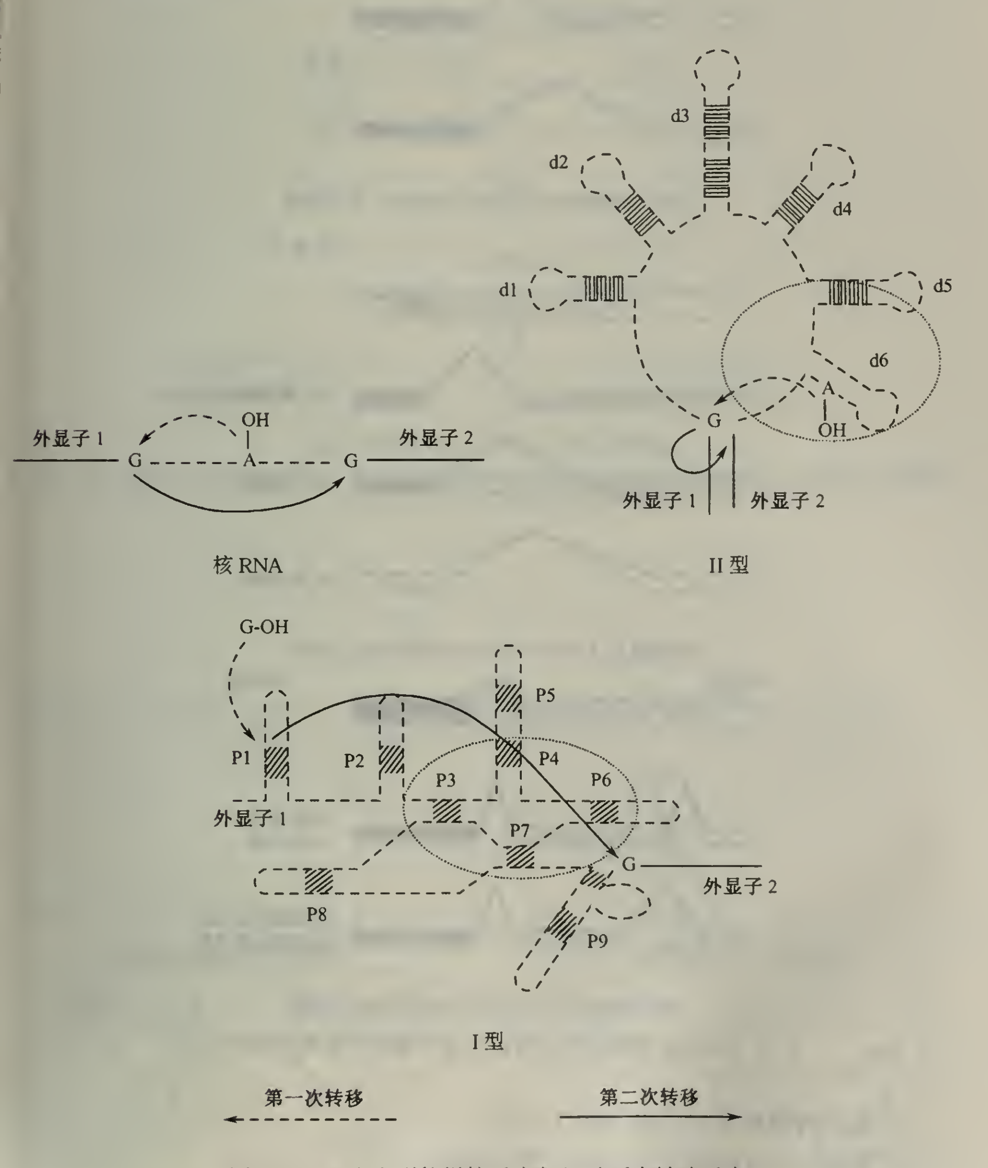
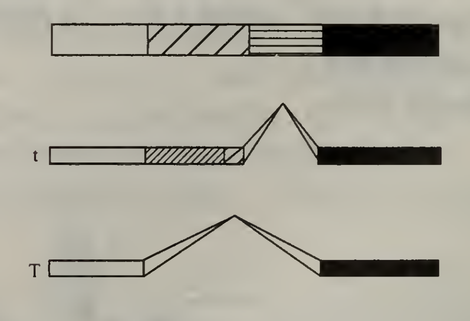
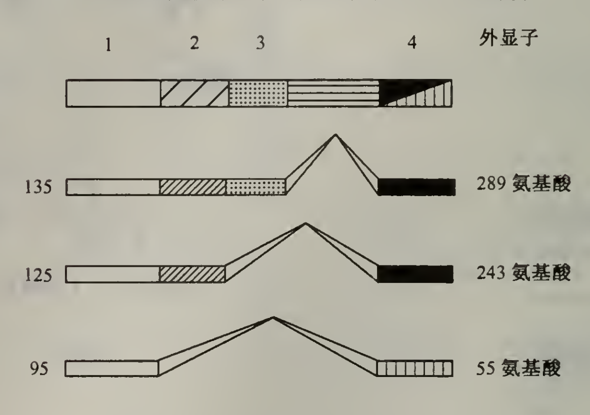


图 7-35　所有类型的拼接反应都经过两次转酯反应

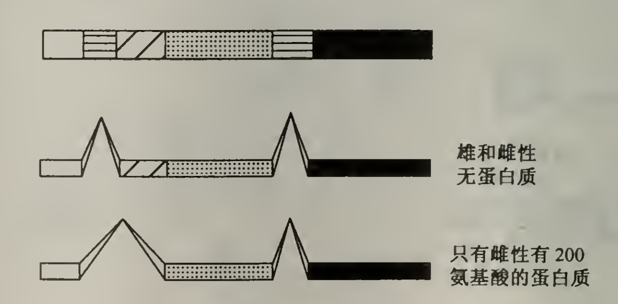
2）选择拼接（图 7-36）；



SV40T/t 抗原两个 5′位点与共同的 3′位点拼接



腺病毒 E1A 多个 5′位点与共同的 3′位点拼接



黑腹果蝇通过 5′位点对不同的 3′位点拼接

图 7-36　改变拼接方式可以从单个基因产生多种蛋白质

3）顺式拼接和反式拼接（图 7-37）；

4）酵母 tRNA 拼接（图 7-38，图 7-39）。

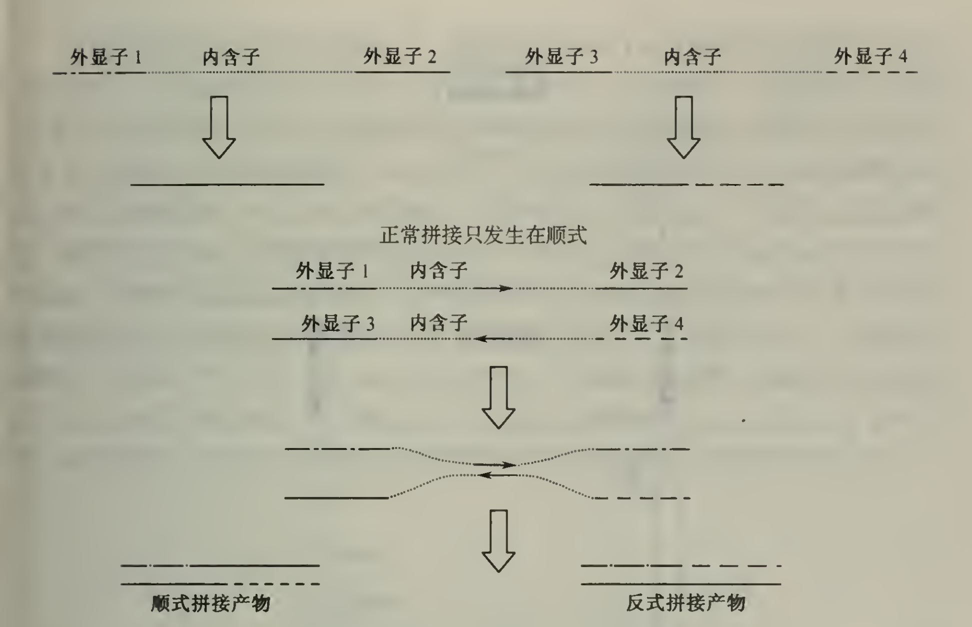


图 7-37　拼接一般以顺式进行，但当造成特殊结构提供内含子碱基配对时，可发生反式拼接

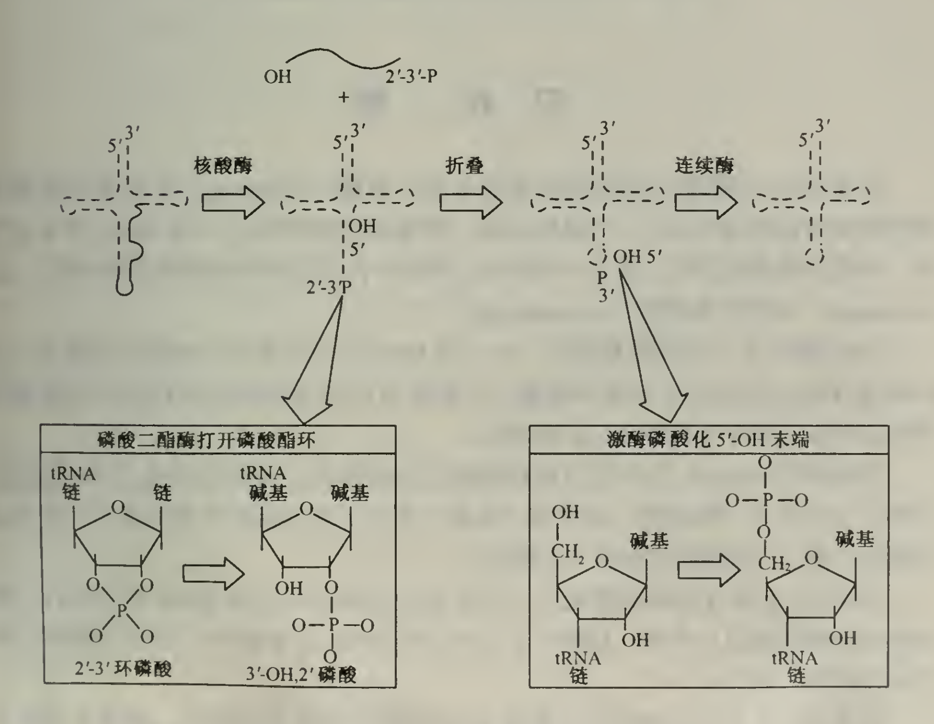


图 7-38　酵母的 tRNA 在体外的拼接可通过凝胶电泳来分析前体 tRNA 和拼接产物

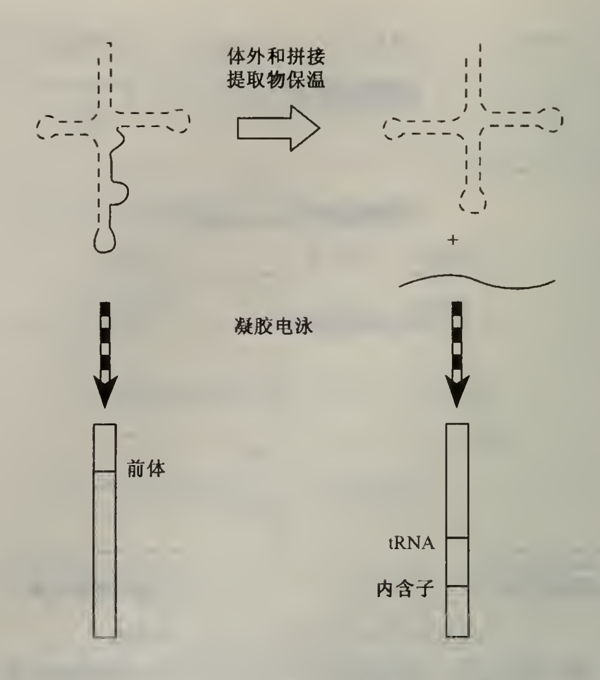


图 7-39　tRNA 的拼接需要核酸酶和连接酶活性

四、核　酶

已知 RNA 中存在几种类型的催化反应，核酶（ribozyme）已成为用来描述具有催化活性的 RNA 的一个通用术语。根据反应的类型，一些 RNA 对非自身的一些底物有催化活性，而另一些 RNA 则在分子内发挥作用称作自我切割（autocleavage）或者自我拼接（autosplicing）。

RNA 酶 P 是一种核糖核蛋白，由一条 RNA 分子与其结合的蛋白质组成，该 RNA 在 tRNA 的加工中具有切割能力，而蛋白质成分起间接的作用，可能用来维持这种具有催化活性的 RNA 的结构。

类病毒类小 RNA 具有进行自我切割反应的能力，该反应发生在分子内部，这种分子可分为“酶活性”部分和“底物”部分，并且相关序列的基因工程可以创造出“酶”，这种酶不依赖于“底物”。

I 类内含子具有自我拼接能力，这种反应在体外可以由 RNA 单独执行，但在体内需要辅助蛋白帮助，I 类内含子的工程 RNA 分子具有与原始功能相关的其他几种催化活性。

在体外，这些反应的共同主体是 RNA 执行切割和通过磷酸二酯键连接的反应。

自我拼接是在四膜虫细胞核内、真菌线粒体和 T4 噬菌体中发现的 I 类内含子的性质。反应需要的信息存在于内含子序列中，反应需要一个特别的二级结构(可能是三级结构)。内含子可能编码负责移动的内切核酸酶。I 类和 II 类内含子均含有可读框。RNA 编辑是在 mRNA 水平上改变信息的过程，RNA 编辑在转录后或转录期间改变一个 RNA 的序列，这种变化可以创造一个有意义的编码序列，编辑过程是通过核酶在转录后或转录中的 RNA 序列中增加或缺失或替换一个碱基，而改变 mRNA 的信息。这种编辑过程在哺乳动物系统中常常发生单个碱基的替换，而在椎虫和副黏病毒中常常发生缺失（常常缺失 U)。在椎虫中，编辑反应非常广泛，有一种 mRNA 中多达一半碱基被编辑。编辑反应使用与 mRNA 互补的导向 RNA 作为模板，导向 RNA 和预编辑（pre-edited）RNA 配对，并利用其 3′端多聚 U 尾巴通过转酯反应把尿嘧啶插入到目标 RNA 中。

第八章　翻　　译

从 20 世纪 50 年代初期开始，经过了 20 多年的努力，已经逐渐了解清楚了蛋白质合成与 DNA、RNA 及核糖体的关系。近年来的研究对象已从原核生物转到真核生物及其细胞器，由整体或复杂体系进入分子水平，许多参与蛋白质合成的生物大分子的一级结构、部分高级结构已被阐明，蛋白质合成的调控问题也取得了不少研究成果。

蛋白质的生物合成过程也就是生物体内遗传信息的“翻译”过程。这一过程比“复制”和“转录”都要复杂得多，它大约需要 200 多种生物大分子，其中包括核糖体（由 RNA 和蛋白质组成）、mRNA、tRNA 以及同工受体 tRNA、氨酰 tRNA 合成酶、可溶性蛋白质因子（起始因子、延伸因子和释放因子）等，参加协同作用。关于蛋白质的生物合成需要论述两方面的问题，一方面是有关遗传信息的传递问题，另一方面就是关于蛋白质合成的机制问题。

基因的遗传信息是通过 mRNA 传递到基因产物——多肽的。DNA 及其转录产物 mRNA 的核苷酸顺序与蛋白质的氨基酸顺序间的关系通过遗传密码来体现。遗传密码是由核苷酸组成的三联体。翻译时，从起始密码子开始，沿着 mRNA 的 5′→3′方向不重叠地连续阅读氨基酸密码子，一直到终止密码子停止，从 N 端到 C 端生成一条具有特定顺序的肽链。

关于遗传信息如何从 DNA 到 mRNA 在前几章里已论述，本章主要讨论遗传密码的破译和蛋白质生物合成的机制。

第一节　遗传密码的破译

遗传密码（code）是指 DNA 上的核苷酸与蛋白质链上的氨基酸之间的对应关系。

密码子（codon）是指 mRNA 上三个连续的核苷酸决定一个特定的氨基酸。mRNA 上特定的密码子顺序与相应蛋白质的氨基酸顺序相对应。

在 Watson 和 Crick 1953 年提出 DNA 双螺旋模型后的第 2 年，物理学家 G. Gamow 首先提出了遗传密码的思想。当时生化遗传学中已有很多实验根据证明“一个基因一个蛋白质”的假说，因此遗传密码的概念易于人们接受。但当时还没有发现 mRNA 和 tRNA，所以对遗传密码的研究很难开展。在相当长的一段时间内，密码破译工作都没有成功。

一、Crick 的探索

1961 年《基因的调控》一文发表，提出了 mRNA 的存在，并能携带遗传信息。以后 Crick 以 T2 噬菌体为材料，用遗传学方法证明了三个核苷酸决定一个氨基酸。他们用吖啶类药物处理大肠杆菌 T2 噬菌体，引起 T2 突变，同时发现这类突变可分为两大类，一类是“+”，另一类是“-”，实际上“+”就是插入了一个核苷酸，使阅读框前移一格，“-”就是减少了一个核苷酸，使阅读框后移一格。他们用这种方法得到了许多突变体，再用这些突变体进行重组实验，如果两个“+”突变株同时感染细胞则得不到野生型重组子，若将两个“-”突变株做重组实验也得到同样结果。但如将三个“+”突变或三个“-”突变做重组实验，均可得到野生型重组子。

上述结果表明，当三个“+”突变或三个“-”突变重组于同一染色体上时，则能恢复野生型，而其他情况如两“-”、两“+”或两“-”、一“+”，两“+”、一“-”都不能重组产生野生型。这个实验初步证明了三个核苷酸决定一个氨基酸。

遗传密码是每三个核苷酸一读，叫做可读框（reading frame)，上述突变实际上就是移框突变。

二、Nirenberg 的实验

1961 年 F. Jacob 提出 mRNA 概念，并与 S. Brenner 等作了实验证实，之后 M. W. Nirenberg 等就开始用实验的方法来研究遗传密码，所用的方法是人工合成一定的 RNA 作为 mRNA，用大肠杆菌的核糖体和其他可溶性提取物及 ATP 做细胞外蛋白质合成试验。最初他们用多核苷酸磷酸化酶来合成 RNA，这种酶能把一种或几种二磷酸核苷酸聚合成多核苷酸，不需要模板，合成的多核苷酸顺序是随机的。例如用 1 种二磷酸核苷酸 PPU 聚合而成的同质多聚体(homopolymer)是多聚 U。以此多聚 U 作为人工 mRNA，进行细胞外蛋白质合成试验。每次试验时用^{14}C 标记 20 种氨基酸中的一种，看哪一种氨基酸能掺入核糖体（离心时随核糖体一起沉降)，结果发现只有苯丙氨酸能掺入。因此知道苯丙氨酸的密码是 UUU，这是第一个译出的密码。以后又译出 AAA 是赖氨酸的密码，CCC 是脯氨酸密码。

1964 年 M. W.Nirenberg 等已能合成一定顺序的二核苷酸和三核苷酸。这个方法是用二环已碳二亚胺（dicyclochexyuar bodimid DCC）把 3′磷酸核苷酸的 3′磷酸和另一核苷酸的 5′-OH 结合：

同样用上述无细胞体系，加入某二核苷酸或三核苷酸，然后看标记氨基酸的

(B₁、B₂ 是两个碱基，缩合时 B_2 的 5′-OH、2′-OH 和 B_1 的 3′-OH 都要保护)

结合情况，实际上就等于在核糖体上加一个只有两个或三个核苷酸的微型 mRNA。结果发现二核苷酸完全不能结合氨基酸而三核苷酸可以，但 UAA、UAG、UGA 3 种三核苷酸不能结合氨基酸为“无字义密码子”(终止密码子)。

三、Khorana 的实验

用磷酸核苷酸合成多核苷酸比较困难，但用磷酸脱氧核糖核苷酸合成 DNA 时由于 2′位上没有—OH 需要保护，因而在技术上容易得到。Khorana 等先合成各种二、三或四核苷酸，再用 DCC 把同一种寡核苷酸缩合成周期性的多核苷酸，例如把两个 AAC 缩合。

(R_1 和 R_2 为保护基团)

如此可以合成 $(AAC)_n$，再合成其互补链 $(GTT)_n$，组成双链 DNA，然后用 RNA 聚合酶作细胞外转录。可得 $(AAC)_n$ 和 $(GUU)_n$ 两种 RNA，用此人工合成的 RNA 作 mRNA 进行细胞外蛋白质合成实验时就不仅是一个 tRNA 分子带一个氨基酸结合到核糖体上，而是在有 GTP 的条件下可以合成多肽链了。其实验结果如表 8-1。

由同种核苷酸聚合而成的周期性多核苷酸所合成的多肽都是只由一种氨基酸所组成的同质多聚体，其中前面 6 种都合成 3 种同质多肽，如 $(AAC)_n$ 合成多苏氨酸或多谷酰胺，此现象说明密码子由 3 个核苷酸组成。它之所以合成 3 种不同的同质氨基酸多聚体是由于阅读框不同而有 3 种翻译方法：

表 8-1 人工合成的 mRNA 及细胞外合成的多肽

人工合成 mRNA	细胞外合成的多肽
$(UC)_n$	$(Ser\text{-}Leu)_n$
$(UG)_n$	$(Val\text{-}Cys)_n$
$(AC)_n$	$(Thr\text{-}His)_n$
$(AG)_n$	$(Arg\text{-}Glu)_n$
$(UUC)_n$	$(Ser\text{-}Ser)_n$，$(Leu\text{-}Leu)_n$，$(Phe\text{-}Phe)_n$
$(UUG)_n$	$(Leu\text{-}Leu)_n$，$(Val\text{-}Val)_n$，$(Cys\text{-}Cys)_n$
$(AAG)_n$	$(Arg\text{-}Arg)_n$，$(Lys\text{-}Lys)_n$，$(Glu\text{-}Glu)_n$
$(CAA)_n$	$(Thr\text{-}Thr)_n$，$(Asn)_n$，$(Gln\text{-}Gln)_n$
$(UAC)_n$	$(Thr\text{-}Thr)_n$，$(Leu\text{-}Leu)_n$，$(Tyr\text{-}Tyr)_n$
$(AUC)_n$	$(Ile\text{-}Ile)_n$，$(Ser\text{-}Ser)_n$，$(His\text{-}His)_n$
$(GUA)_n$	$(Ser\text{-}Ser)_n$，$(Val\text{-}Val)_n$
$(GAU)_n$	$(Asp\text{-}Asp)_n$，$(Met\text{-}Met)_n$
$(UAUC)_n$	$(Tyr\text{-}Leu\text{-}Ser\text{-}Ile)_n$
$(UUAC)_n$	$(Leu\text{-}Leu\text{-}Thr\text{-}Thr)_n$
$(GAUA)_n$	不合成多肽
$(GUAA)_n$	不合成多肽

CAA，CAA，CAA，……

AAC，AAC，AAC，……

ACA，ACA，ACA，……

这三种密码子必定分别代表苏氨酸（Thr）、天门冬酰胺（Asn）、谷氨酰胺（Gln），但是还不能确定哪一个密码子代表哪种氨基酸。

由二核苷酸聚合而成的多核苷酸共四种，每种都合成一种多肽，而且都由两种氨基酸周期性交替而成，例如（AC）$_n$ 所合成的多肽是 Thr-His-Thr-His……，这是因为（AC）$_n$ 只有一种读法，即：ACA，CAC，ACA，CAC……。

因此 ACA 和 CAC 必定分别代表 Thr 和 His，但也不能确定哪一个密码代表哪一个氨基酸。

如果把以上两个结果合起来看，就知道 ACA 一定是编码 Thr，因此 CAC 一定编码 His。

需要注意的是，这是在特定条件下的细胞外翻译，不需要 AUG 起始密码子，因此阅读框不固定，所以才可能有这个结果。

综合以上 3 个小组的实验结果和其他实验室用比较间接的方法所得的结果，

到 1966 年遗传密码全部破译，得出密码表（表 8-2）。

表 8-2　遗传密码表

第一个字母 5′端	第二个字母 U	第二个字母 C	第二个字母 A	第二个字母 G	第三个字母 3′端
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	U
	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	C
	UUA Leu	UCA Ser	UAA Stop	UGA Stop	A
	UUG Leu	UCG Ser	UAG Stop	UGG Trp	G
C	CUU Leu	CCU Pro	CAU His	CGU Arg	U
	CUC Leu	CCC Pro	CAC His	CGC Arg	C
	CUA Leu	CCA Pro	CAA Gln	CGA Arg	A
	CUG Leu	CCG Pro	CAG Gln	CGG Arg	G
A	AUU ILe	ACU Thr	AAU Asn	AGU Ser	U
	AUC ILe	ACC Thr	AAC Asn	AGC Ser	C
	AUA ILe	ACA Thr	AAA Lys	AGA Arg	A
	AUG Met	ACG Thr	AAG Lys	AGG Arg	G
G	GUU Val	GCU Ala	GAU Asp	GGU Gly	U
	GUC Val	GCC Ala	GAC Asp	GGC Gly	C
	GUA Val	GCA Ala	GAA Glu	GGA Gly	A
	GUG Val	GCG Ala	GAG Glu	GGG Gly	G

后来许多研究结果都证明这个密码表是正确的，特别是在测定天然 DNA 和 RNA 中核酸顺序时，同时测定其所确定的氨基酸顺序，如 *ΦX*174 及 MS2 等，结果都与密码表完全相符。

随着对遗传密码的进一步研究，发现它们具有一些重要性质。

1．简并与兼职

一共有 64 个密码子，除 3 个终止密码子外，余下 61 个为氨基酸密码子，而氨基酸有 20 种，所以许多氨基酸的对应密码子不止 1 种，这种现象叫做密码的简并（degenerate）。对应于同一氨基酸的不同密码子称为同义密码子（synonym codon）。

在已知的 61 个氨基酸密码子中，有两个是兼职的，它们除了代表特定的氨基酸外，还兼作起始密码子，作为肽链合成的起始信号，这两个密码子是 AUG（Met）和 GUG（Val）。AUG 使用得最为普遍。

2．密码子中不同位置碱基的特异性

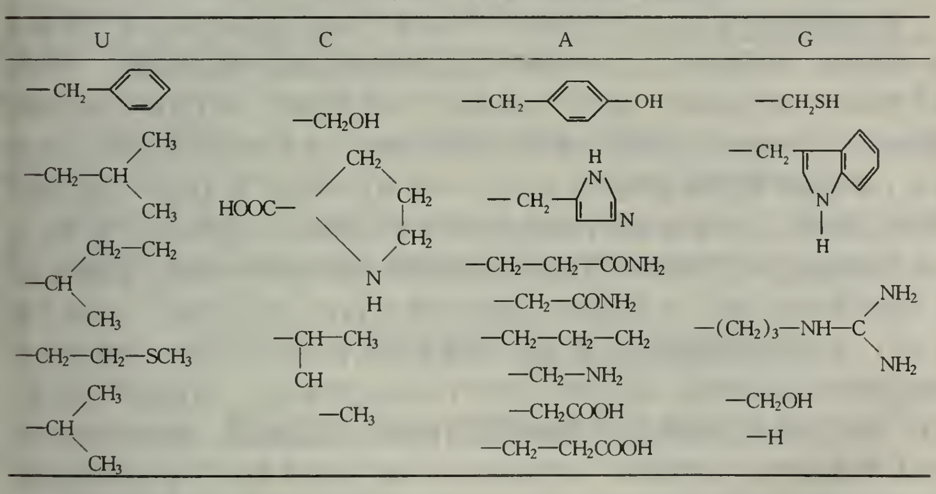
虽然密码子的简并程度各有不同，但同义密码子在密码子字典中不是杂乱分

布的。64 个密码子被分配在 16 个方格内，凡属同一氨基酸的密码子，都位于同一方格内，这些密码子的前两个碱基相同，区别只在于第 3 个碱基：或为任一碱基（如 Pro、Thr、Ala、Gly 等密码子）；或同是嘌呤（如 Gln、Glu、Lys 密码子）；或都是嘧啶（如 Phe、Tyr、Cys，His、Asn 和 Asp 密码子）。这表明密码表中前两个碱基的特异性很强，有 6 个密码子的氨基酸（Leu、Arg 和 Ser）也是以方格区分的。

Met 和 Trp 都只有 1 个密码子，Ile 有 3 个密码子，似乎与其他氨基酸密码子有所不同。但是，近年来研究哺乳动物线粒体密码子使用情况时发现，这 3 个有奇数密码子的氨基酸均变为含偶数密码子的氨基酸，而且也符合密码子中前两个碱基特异性强的规律。

在密码子的 3 个碱基中，有人认为第 2 个（中间的）碱基与氨基酸的极性有关：若氨基酸具极性侧链，密码子的第 2 个碱基是嘌呤；若侧链为非极性，则第 2 个碱基是嘧啶（也有个别例外），见表 8-3。

表 8-3　密码子中第 2 个碱基与氨基酸侧链的关系



U	C	A	G
$—CH_2—C_6H_5$（苯环）	$—CH_2OH$	$—CH_2—C_6H_4—OH$	$—CH_2SH$
$—CH_2—CH(CH_3)_2$	$HOOC—$（吡咯烷环，CH_2、CH_2、CH_2、NH）	$—CH_2—$（咪唑环，NH、N）	$—CH_2—$（吲哚环，N—H）
$—CH(CH_2—CH_2)(CH_3)$	$—CH(CH_3)—CH$	$—CH_2—CH_2—CONH_2$	$—(CH_2)_3—NH—C(NH_2)NH_2$
$—CH_2—CH_2—SCH_3$	$—CH_3$	$—CH_2—CONH_2$	$—CH_2OH$
$—CH(CH_3)_2$		$—CH_2—CH_2—CH_2$	$—H$
		$—CH_2—NH_2$	
		$—CH_2COOH$	
		$—CH_2—CH_2COOH$	

3. 通用与例外

密码子的一个最重要的性质是其通用性，无论体内、体外实验，从病毒、原核到真核，各种不同生物都可使用同一密码子字典。

但是，1980 年以来通过对人、牛和酵母线粒体基因序列和基因结构的研究发现，密码子的通用性也有例外，在这些基因中并不完全按照密码子字典编码。以人和牛线粒体为例，有以下改变：①AUA 是 Met 而不是 Ile 的密码子。这样，Ile 由原来的 3 个密码子改为两个，而 Met 由原来的 1 个增至两个，变动后出现

两个新家族：AUU 或 AUC（Ile）和 AUA 或 AUG（Met）。②UGA 不是终止密码子，而成为 Trp 密码子。结果 Trp 密码子也由一个增至两个，成为一个新家族：UGA 或 UGG（Trp）。③AGA 和 AGG 不是 Arg 密码子，而变为终止密码子，在哺乳动物线粒体中共有 4 个终止密码子，分为两个家族：UGA 或 UGG 和 AGA 或 AGG。上述有异于密码子字典的规律的改变，使所有密码子都成为偶数，全部都可以按家族区分。这表明仍有规律可循，其意义有待进一步探讨。④除AUG 外，尚可兼用 AUA 和 AUU 作为起始密码子。

4．密码子的阅读方向与 mRNA 的编码方向一致

在蛋白质生物合成过程中，mRNA 的编码方向是 5′→3′，从 N 端向 C 端延伸肽链，因此密码子阅读方向也是从 5′端至 3′端。一条肽链的合成起始之后，密码子按 3 个一框读下去，不重叠也不跳格，中间也没有标点（即不停顿）直到终止。

5．密码子的使用情况

由于核酸特别是 DNA 序列分析的飞快发展，目前已测出了许多蛋白质基因的一级结构。将基因或其 mRNA 序列与蛋白质的氨基酸序列进行比较，不仅验证了密码子字典的正确性，提供了许多密码子的使用情况，并且发现了新的氨基酸密码子。①起始密码子和终止密码子的使用情况：绝大部分使用 AUG 为起始信号（有报道真核生物全部使用 AUG），个别使用 GUG；而 UAA、UAG 都可作为终止密码子。在某些基因中有时连用两个终止密码子，可能是为了更为有效地终止肽链的合成。②几种哺乳动物氨基酸密码子使用频率的比较：对鸡卵白蛋白（密码子数为 388）、兔 β-球蛋白（密码子数为 147）、兔 α-球蛋白（密码子数为 142）、免疫球蛋（密码子数为 128）和胰岛素（密码子数为 99）的密码子使用情况的研究结果表明，它们对同义密码子的使用差异甚大，有的被广泛使用，而另一些则使用得比较少。③大肠杆菌基因密码子的使用情况：原核生物的代表大肠杆菌基因的研究开展最早，进展也最快。通过比较大肠杆菌中 13 种含量较多的蛋白质密码子的使用情况发现这些高表达的基因中密码子的使用有以下规律：对于以 C 或 U 结尾的同义密码子，如前两个碱基为 A、U，则第三个碱基优先使用 C 而不是 U；若前两个碱基为 C、G，则第三个碱基优先使用 U 而不是 C。这样的选择有利于翻译时密码子与反密码子的相互作用，两者的作用能不太强也不太弱，而以中度为宜，即所谓最适相互作用能。MS2 噬菌体密码子的使用情况与大肠杆菌基因密码子的使用有相似的规律。④哺乳动物线粒体基因的密码子使用情况：人和牛线粒体基因编码 13 种蛋白质，如前所述，线粒体使用的起始和终止密码子以及 3 个氨基酸（Met、Trp 和 Ile）的密码子与其他生物不同。线粒体的这种遗传体系既不同于原核生物又区别于真核生物，人和牛线粒体

基因在使用同族密码子中，较多使用的是以 A 或 C 结尾的密码子。

酵母线粒体密码子中，较多使用的情况既不同于胞质，也不同于哺乳动物线粒体。

第二节　摇摆假设

为什么不同密码子代表同一氨基酸？特别是密码子的第三个字母容许有变化，Gly、Pro、Thr、Val、Ala 中，似乎只由密码子的前面两个字母决定，第三个字母是 U、C、A、G 都可以。前面两个字母决定之后，第三个字母是 U 或 C 往往意义相同，如 Cys、Asn、Asp、His、Phe；第三个字母是 A 或 G 的意义往往也相同，如 Gln，Glu、Lys 等。再加上 *E.coli* 中并不存在 61 种以上的 tRNA 的事实，促使 F.H.C.Crick 在 1966 年提出摇摆假设。他根据分子模型假设：tRNA 的反密码子的第一个字母如果是 U、φ、G 或 I 的话，则可以和 mRNA 密码子第三位上一种以上的碱基配对，如表 8-4。

表 8-4　tRNA 反密码子与 mRNA 密码子的配对

tRNA 反密码子第一碱基	mRNA 密码子第三碱基
C	G
A	U
U 或 φ	A 或 U
G	C 或 U
I	U，G 或 A

G 在摆动键角时，可以和 U 形成两个氢键，因此像 Cys、Asn 或 Gln、Glu 等氨基酸只要一种 tRNA 就够了，像 Ala、Gly、Pro 等也只要两种 tRNA。

摇摆假设后来为许多实验证明，因为凡是已提纯的 tRNA 都已测定核苷酸顺序，其反密码子是已知的，可以用细胞外实验直接测知其识别密码子的序列。

第三节　蛋白质生物合成的机制

一、与蛋白质生物合成有关的生物大分子

蛋白质生物合成是由许多种类的生物大分子协同作用进行的，因此在讨论合成机制之前有必要分别论述这些分子的结构、功能及性质。

1. 核糖体及多核糖体

核糖体是由几十种蛋白质和几种 RNA 组成的一种亚细胞颗粒体，它是蛋白质生物合成的场所。

核糖体中蛋白质占其重量的 1/3，RNA 占 2/3。经过 20 多年研究，对它们的种类及其在亚基上的分布已经有了一个清楚的轮廓。从表 8-5 可以看出，真核生物相应的 RNA 比原核的沉降系数大，蛋白质的数目比较多；而线粒体核糖体

的 RNA 的沉降系数一般比原核生物的小，但蛋白质数目一般比原核生物的多；真菌和动物线粒体的核糖体中尚没有发现 5S RNA 和 5.8S RNA；叶绿体的情况则与原核生物非常相似，但大亚基中有三种 rRNA。

表 8-5　核糖体中的 RNA 和蛋白质

不同来源	核糖体/S	亚基/S	RNA/S	蛋白质种类
真核生物[①]	80	40	18	30～32
		60	28，5.8，5	36～50
原核生物[②]	70	30	16	21
		50	23，5	34
线粒体	64～70	32～40	11～16	30～44
		43～58	14～23	31～50
叶绿体[③]	68～70	28～30	16	
		33～55	23，5，4，5[③]	

①细胞质；②大肠杆菌；③存在于高等植物叶绿体中。

对大肠杆菌核糖体的组分和化学结构已全部研究清楚：小亚基由一种 RNA（16S，1542 个核苷酸）和 21 种蛋白质构成；大亚基则由两种 RNA（23S，2904 个核苷酸和 5S，120 个核苷酸）和 34 种蛋白质构成。

大肠杆菌核糖体小亚基共有 21 种蛋白质，分别用 S1 至 S21 表示；大亚基有 34 种，表示为 L1 至 L34。两亚基间有两个蛋白质的结构完全相同，即 S20 = L26；所以实际上大肠杆菌核糖体蛋白质只有 54 种而不是 55 种。另外 L7 和 L12 的氨基酸数目完全相同，只是 L7 N 端为乙酰丝氨酸而 L12 是丝氨酸。至于核糖体中各种蛋白质的数目，除 L7/L12 为四分子外，其他都是以单分子存在。有 70％以上的蛋白质为碱性蛋白质，等电点在 pH 10 以上，只有三种蛋白质 S6、L7 和 L12 的等电点在 pH 5 左右。有几种蛋白质含有修饰氨基酸，如 N 端氨基酸被乙酰化，以及内部氨基酸的甲基化，个别蛋白质如 L11 有 9 个甲基。核糖体是一个高度复杂的体系，任一组分或局部组分都不能单独起到整体作用。

用重组技术、化学交联和化学修饰技术、荧光标记技术、中子衍射法、免疫电镜法和亲和层析法对核糖体的空间拓扑学进行了广泛的研究。用上述方法配合大量的实验，得到了大量关于 rRNA 之间以及 rRNA 和核糖体蛋白之间关系的结果。用共价交联法分离出了 RNA 与蛋白质的复合物，如 L1、L5、L18 和 L25 与大亚基中的 RNA 有关，L1 与 23S rRNA 中一段通用序列（56 个核苷酸）结合，5S rRNA 和 23S rRNA 则通过 L5、L13 而相互作用。通过对 RNA 与蛋白质结合的初步研究表明，在 30S 亚基中两者的分布是均匀的，大约有 2/3 的蛋白质与 RNA 直接结合；而在 50S 亚基中，两者的分布是不均匀的，即蛋白质集中于一

端而 RNA 集中于另一端，蛋白质中心与 RNA 中心距离约 50～60Å（中子衍射结果及免疫电镜法支持这一结论）。两个亚基中蛋白质的分布差别这么大，可能反映它们在功能上不同。

随着电镜技术的发展，获得多种角度的清晰的二维结构图像，再用数学推算出它与一个晶体三维结构的关系；根据这些结果，Lake 提出大肠杆菌核糖体模型，见图 8-1。小亚基包括一个头部，一个基部和一个平台，平台与头部之间有一裂口；大亚基有一个中心突出部分，一边有一个脊，另一边有一个柄。结合上述数据，核糖体组分特别是蛋白质在核糖体中的拓扑位置基本上可以确定。

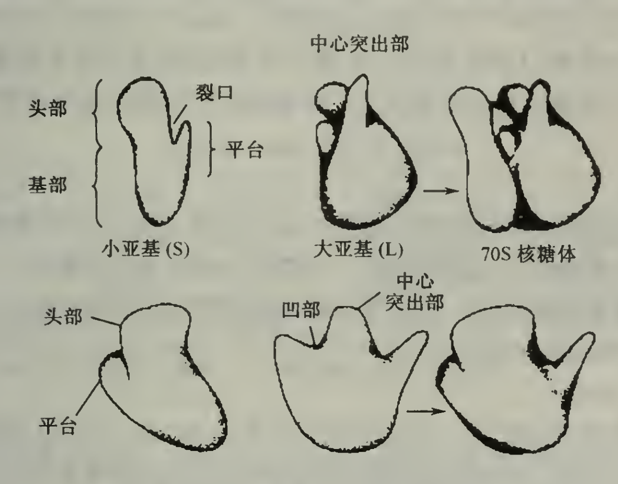


图 8-1　核糖体的三维结构模型

无论原核生物还是真核生物，从中都可同时分离出三种类型核糖体：核糖体、核糖体亚基和多核糖体，它们之间的关系见图 8-2；在体内蛋白质合成正是通过这种核糖体循环进行的。在细胞中，大多数核糖体处于非活性的稳定状态，单独存在；只有少数与 mRNA 形成多核糖体。在多核糖体中，每个核糖体单独

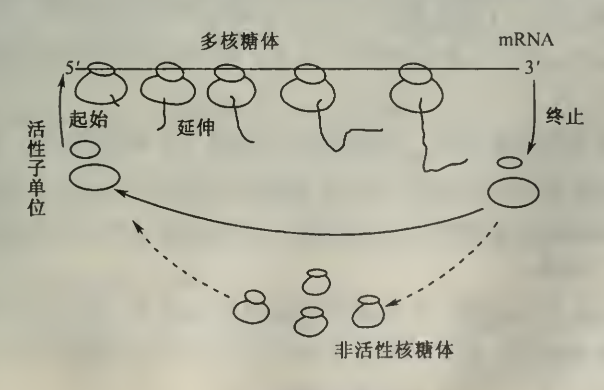


图 8-2　核糖体循环图

合成一条完整的多肽链。mRNA 上核糖体的多少，视 mRNA 链长而定，一般每隔 40 个核苷酸有一个核糖体。新生肽链释放后，核糖体即脱离 mRNA，很快解聚成核糖体亚基。这些亚基有两个去向，一个去向是直接参与另一轮蛋白质合成；另一个去向是两个亚基结合成较稳定的核糖体，稳定核糖体不参加蛋白质合成，只有在解聚为亚基后才能再成为活性核糖体亚基。

真核生物核糖体，有的游离于细胞质中，有的结合于内质网上。游离态与结合态的比例因细胞类型不同而异。

2. mRNA

mRNA是合成蛋白质的模板，由蛋白质基因转录、加工而得到。成熟的 mRNA 中，除了编码区还有非编码区，非编码区对于翻译也是必要的。

3. tRNA

tRNA 是蛋白质生物合成中的运载工具，根据 mRNA 序列编码的需要携带相应氨基酸进入核糖体；合成肽链后，“空的” tRNA 离开核糖体。

关于 tRNA 的生物合成、结构及功能生物化学中已详细讨论过，这里着重介绍校正 tRNA 和起始 tRNA。

（1）校正 tRNA

我们在无义突变的校正一节已经介绍了校正 tRNA，实际上就是编码 tRNA 的基因的另一种突变。这些突变使转录出的 tRNA 的个别碱基发生改变，尤其是反密码子环上一个碱基的改变，使原来与某一氨基酸的密码子对应的反密码子能够和终止密码子对应；因而将某一氨基酸插到该密码子处，使肽链合成得以延伸。

对校正 $\text{tRNA}^{\text{Tyr}}_{\text{Su3}^{+}}$ 的一级结构研究表明，只有反密码子的第一位碱基由 G 变为 C，其他部分与 $\text{tRNA}^{\text{Tyr}}_{\text{Su3}^{-}}$ 完全相同。这个碱基的变化完全改变了 tRNA 识别密码子的能力。$\text{tRNA}^{\text{Tyr}}_{\text{Su3}^{-}}$（其反密码子 GUA）能识别酪氨酸的两个密码子 UAC 和 UAU，而校正 $\text{tRNA}^{\text{Tyr}}_{\text{Su3}^{+}}$（其反密码子 CUA）就只能识别变异产生的终止密码子 UAG。

$\text{tRNA}^{\text{Tyr}}_{\text{Su3}^{+}}$ 再发生一次变异，反密码子以外的一个碱基被另一个碱基所取代，产生的校正 tRNA 仍可以识别由变异产生的终止密码子 UAG，但携带的不是酪氨酸而是谷氨酰胺了。这种校正 $\text{tRNA}^{\text{Gln}}_{\text{Su3}^{+}}$ 已发现有五种，分别在第 1、2、81、82 位变化了一个碱基。

在校正 tRNA 中，即使一个碱基的变化就可以改变 tRNA 识别密码子的能力或者接受不同的氨基酸。研究这类校正 tRNA 的结构与功能的关系，对于探索氨酰 tRNA 合成酶的专一性问题有着重要的意义。

（2）起始 tRNA

能特异识别 mRNA 上起始密码子的 tRNA 称为起始 tRNA（initiator tRNA），它参与蛋白质生物合成的起始。所有其他的 tRNA 在蛋白质合成中起延伸作用，统称为延伸 tRNA。已知起始密码子有两个——AUG 与 GUG，它们分别是 Met 和 Val 的密码子。核酸序列分析结果指出，除了少数生物使用 GUG 作为起始密码子外，大部分原核生物、病毒及差不多全部真核生物都使用 AUG。也就是说起始 tRNA 都携带 Met，核糖体上合成的新链的 N 端均为 Met。原核生物与真核生物又有所不同，原核生物合成的蛋白质前体中 N 端 Met 的 α-氨基被甲酰基所封闭，得到甲酰甲硫氨酸（fMet）。故起始 tRNA 是一种特殊功能的甲硫氨酸 tRNA，它们的符号见表 8-6。

表 8-6　起始 tRNA 及其符号

tRNA 种类	来　源	tRNA	氨酰 tRNA	功　能
*起始 tRNA	真核生物	$tRNA_i^{Met}$	$Met\text{-}tRNA_i^{Met}$	真核蛋白合成起始
	原核生物	$tRNA_f^{Met}$	$fMet\text{-}tRNA_f^{Met}$	原核蛋白合成起始
**延伸 $tRNA^{Met}$		$tRNA_m^{Met}$	$Met\text{-}tRNA_{(m)}^{Met}$	肽链内部的 Met 的掺入

* tRNA 右上角的 Met 有时可以省略。** 符号中的 m 字母可以省略。

真核生物起始 tRNA 与 $tRNA_m^{Met}$ 结构上的明显差异在于 AUCG 序列代替通常的 TΨCG 序列。原核生物起始 tRNA 必须先甲酰化后才能行使功能。

（3）氨酰 tRNA 合成酶

氨酰 tRNA 合成酶即氨基酸-tRNA 连接酶，此酶催化下列反应：

$$aa + ATP + tRNA \rightarrow aa\text{-}tRNA + AMP + ppi$$

它使氨基酸的羧基与 tRNA 上 3′端 CCA 中腺苷的 2′或 3′-OH 与氨基酸的羧基形成酯键。不同氨基酸和 tRNA 的连接位置不同，氨基酸密码子第二个字母是 U 的与 tRNA 末端 2′-OH 相连，中间字母是 C 的氨基酸与 3′-OH 连接，中间为 A 或 G 时连接规律性不大。

每一种氨基酸至少有一种氨酰 tRNA 合成酶。可能一个例外是在枯草杆菌中不存在 Gln-tRNA 合成酶。Glu-tRNA 合成酶可以用 Glu 酰化 $tRNA^{Glu}$和 $tRNA^{Gln}$，然后通过转酰氨反应使 Glu-$tRNA^{Gln}$转换成 Gln-$tRNA^{Gln}$。一种氨基酸可以对应多种 tRNA，在氨酰 tRNA 合成酶催化下可以接受同一种氨基酸的不同 tRNA 叫同工受体 tRNA（iso-acceptor tRNA）。

迄今已有 150 种以上不同来源的氨酰 tRNA 合成酶被分离纯化。原核生物中的合成酶一般分子质量≤2.5×10^5Da。而真核生物的酶常以复合形式存在，分子质量＞10^6Da。

4. 起始因子

起始因子是参与蛋白质生物合成起始的可溶性蛋白因子。蛋白质合成的起始要生成核糖体·mRNA·起始 tRNA 三元复合物，也叫起始复合物。复合物必须在起始因子帮助下才能形成。目前已知原核生物有 3 种起始因子，而真核生物则有 10 种。大肠杆菌起始因子的不同形式、分子质量及所起作用见表 8-7。

表 8-7　大肠杆菌的起始因子

起始因子	不同形式	分子质量（Da）	起始中的作用
IF-1		9500	无专一功能，增加 IF-2，IF-3 活性
IF-2	IF-2a	95 000～117 000	使 fMet-tRNAf 选择性与 30S 亚基结合，需
	IF-2b	82 000～85 000	GTP
IF-3	IF-3α	20 668	①使 30S 与 mRNA 起始部位连接
	IF-3β	19 997	②具解离活性，使亚基保持为解离态

IF-1 为一小的碱性蛋白，没有专一的功能，只能增加其他两个起始因子的活性。有的原核生物并没有 IF-1 起始因子，看来蛋白质的合成起始可以只在 IF-2和 IF-3 参与下进行。IF-2 有两种形式：IF-2a 和 IF-2b，后者可能是前者的蛋白酶降解产物。-SH 为 IF-2 活性所必需，IF-2 经磷酸化后活性不变。IF-2 的功能是通过生成 IF-2·GTP·fMet-$tRNA_f^{Met}$ 三元复合物，使起始 tRNA 与核糖体小亚基结合。IF-2 与 fMet-$tRNA_f^{Met}$ 间的专一性是非常严格的，用于延伸的氨酰 tRNA 甚至非甲基化的 Met-$tRNA^{Met}$均不能与 IF-2 结合或结合得很不稳定。在核糖体存在时，IF-2 具有很强的 GTPase 活性，在肽链合成起始时催化 GTP 水解。

IF-3 是个双功能蛋白质，它至少有两种分子形式：IF-3α 和 IF-3β，后者只比前者少 N 端 6 个氨基酸残基。IF-3 除了使 mRNA 与核糖体小亚基结合外，还具解离因子活性。IF-3 与 30S 亚基结合，位于核糖体表面与核糖体蛋白 S7、S12 和 16S rRNA 结合。可能是由于 IF-3 遮盖了小亚基的 16S rRNA 的部分序列使之不能与大亚基的 23S rRNA 互补，故大小亚基不能缔合。

真核生物的起始因子对于生成 80S 核糖体·mRNA·Met-tRNA 三元起始复合物是必要的。真核生物起始因子（eIF）确切的数目仍不清楚，它们是核糖体或 mRNA 结合蛋白。不同作者发表不同来源 eIF 的名称和符号也非常混乱，经统一后列于表 8-8。

eIF-2 是真核生物中最重要的一个起始因子，它与 GTP 和 Met-tRNAi 生成三元复合物（eIF-2·GTP·Met-tRNAi），这个复合物比大肠杆菌的 IF-2·GTP·fMet-tRNAf 稳定。eIF-2 由三个亚基组成，其中 α 亚基与鸟苷酸结合有关，β-亚

表 8-8　真核生物 mRNA 翻译的起始因子

起始因子	分子质量/Da	亚基(分子质量)/Da	已知功能
eIF-1	15 000		稳定 40S·mRNA·Met-rRNAi 复合物
eIF-2	122 000	α(32 000) β(35 000) γ(55 000)	被 GTP 活化，使 Met-tRNAi 与 40S 亚基结合
eIF-2A	65 000		生理功能不详，天然 mRNA 翻译不需要，用于 AUG 指导的合成
eIF-3	700 000	由 9～10 种亚基组成其中至少 7 种的分子质量已测出	与 40S 核糖体亚基结合，也与 mRNA 上结合的起始因子 eIF-4F 互相作用及 40S·Met-tRNAi-GTP 有关
eIF-4A	50 000		辅助 mRNA 结合
eIF-4B	80 000		识别 mRNA，有 ATPase 活性
eIF-4C	19 000		与 40S 结合
eIF-4D	17 000		生理功能不详，天然 mRNA 翻译不需要
eIF-4E	240 000		帽子结合蛋白，与 eIF-4G 相互作用
eIF-5	150 000		释放起始因子，缔合核糖体亚基
eIF-4		eIF-4E eIF-4G	与 mRNA 结合

基有结合 Met-tRNAi 的活性，而 γ 亚基功能不详。已分离出 α、β 两个亚基的 eIF-2，推测 γ 亚基不是必需的。eIF-2 的 β 亚基也与识别 mRNA 有关。

eIF-3 与 *E.coli* 的 IF-3 类似，与 40S 亚基结合，并通过 eIF-4G 的介导使 40S 亚基与 mRNA 结合；同时，eIF-3 的作用可避免 60S 与 40S 核糖体亚基的缔合，因而能逐步生成起始复合物。eIF-3 也具有稳定 40S·Met-tRNAi·GTP 的功能。

参与 mRNA 与 40S 核糖体亚基结合的因子有 eIF-3、eIF-4A、eIF-4B 以及 eIF-1、eIF-4C 等 5 个因子。eIF-4F 与 mRNA 结合，它含 eIF-4E 和 eIF-4G。eIF-4G 是一种十分重要的真核 mRNA 起始因子，在把 40S 亚基招集到 mRNA 上的过程中起桥梁作用；它的 C 端与 eIF-3 互相作用，N 端与 mRNA 结合蛋白如 eIF-4E 作用，还可能与 polyA 结合蛋白互相作用；在这些因子的共同作用下，形成 40S·mRNA·Met-tRNAi 复合物。eIF-4E 是从 eIF-3 和 eIF-4B 中分离的微量成分，它识别 mRNA 的帽子结构，故称帽子结合蛋白。

eIF-5 的功能可能是首先与 40S 核糖体亚基上 eIF-3 和 eIF-2 的释放及 GTP 水解有关，然后缔合 60S 亚基得到起始复合物 80S·mRNA·Met-tRNAi。

eIF-4D 只用于人工信使的翻译，天然 mRNA 的翻译是不需要的，其生理功能不详。

近年来在真核生物中还发现许多与起始反应有关的辅助因子（auxiliary factor）。这些因子大多数与 eIF-2 有关，表 8-9 列出它们的性质与已知功能。Co-eIF-2A、Co-eIF-2B 和 Co-eIF-2C 与 eIF-2·Met-tRNAi·GTP 三元复合物生成的调控有关。eIF-2 激活蛋白（ESP）稳定 eIF-2·GTP 二元复合物，加入 Met-tRNAi 形成三元复合物后，ESP 释放出来。

表 8-9　与蛋白质合成起始有关的辅助因子

辅助因子	分子质量/Da	已知功能
Co-eIF-2A	20 000	稳定三元复合物，增加 Met-tRNAi 与 40S 核糖体亚基的结合速率
Co- eIF-2B	450 000（含 12～15 个肽的复合物）	三元复合物的生成
Co-eIF-2C		促进三元复合物的生成
eIF-2 激活蛋白		增加 eIF-2 生成三元复合物和 40S 起始复合物的能力
eIF-6	23 000～25 500	抗 60S 与 40S 核糖体亚基的缔合

5. 延伸因子

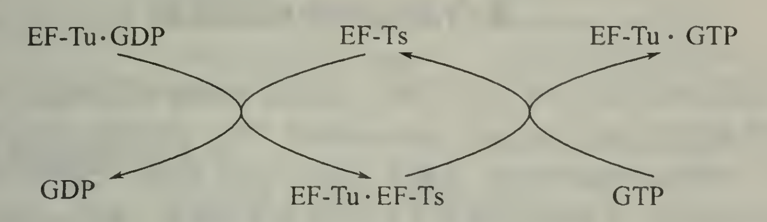
延伸因子是一类参与蛋白质合成过程中肽链延伸的蛋白因子。无论原核还是真核生物，延伸因子都可分为两类：一类是帮助氨酰 tRNA（延伸 tRNA）进入核糖体与 mRNA 结合，另一类是使肽基 tRNA 从核糖体的 A 位向 P 位移动。前者有 EF-T（包括 EF-Tu 和 EF-Ts，细菌中）和 EF-1（真核生物）；后者是 EF-G（细菌）和 EF-2（真核生物）。表 8-10 列出蛋白质合成中的延伸因子。

表 8-10　蛋白质合成的延伸因子

延伸因子	分子质量/Da	—SH 含量
原核生物		
EF-Tu	42 000	3
EF-Ts	31 000	1～2
EF-G	84 000	4～6
真核生物		
EF-1	47 000～60 000	1（兔网织红细胞） 3～4（猪肝）
$EF-1_L$（单体形式）		
$EF-1_H$（聚集形式）		
EF-2	70 000～95 000	18（兔网织红细胞） 15（猪肝） 22（鼠肝）

在原核生物中，EF-Tu 的功能是依据 mRNA 上的密码子携带 aa-tRNA 进入

A 位。EF-Tu 在 GTP 存在下与 aa-tRNA 形成稳定的三元复合物：EF-Tu· GTP· aa-tRNA。EF-Tu 具高度专一性，它能识别和结合除了 fMet-$tRNA_f^{Met}$ 外的所有 aa-tRNA。这样就保证起始 tRNA 携带的 fMet 不能进入肽链内部，因为无论甲酰或非甲酰的起始 tRNA 都不能与 EF-Tu 和 GTP 生成复合物。GTP 水解后，EF-Tu·GDP 从核糖体释放出来。EF-Tu 是一个热不稳定的蛋白因子。EF-Ts 则是热稳定的，它不直接与核糖体发生关系，其功能是使 EF-Tu·GDP 再生为 EF-Tu·GTP，然后再参与肽链延伸：



在蛋白质合成的延伸机制中有一个移位过程，即肽基 tRNA 从 A 位移至 P 位，核糖体与 mRNA 相对移动 3 个核苷酸（1 个密码子）距离，这个复杂过程需要 EF-G 和 GTP。在核糖体存在下，EF-G 使 GTP 水解，故 EF-G 是依赖于核糖体的 GTPase。

真核生物的延伸因子是 EF-1，与原核生物的 EF-Tu 类似。EF-1 可以是单体形式（$EF\text{-}1_L$），也可以是各种聚集态（$EF\text{-}1_H$）；$EF\text{-}1_H$ 是 EF-1 通过其他多肽、胆固醇酯聚合起来的。$EF\text{-}1_H$ 类似于原核生物的 EF-Ts，启动 EF-1 循环。

EF-2 类似于原核生物的 EF-G，为肽基 tRNA 从 A 位移位所必需。EF-2 也具有依赖于核糖体的 GTPase 活性，GTP 水解后，EF-2 从核糖体释放出来。

6. 释放因子（终止因子）

原核生物中发现有三种释放因子（release factor）——RF-1、RF-2 和 RF-3，哺乳动物只有一种释放因子 RF。释放因子的作用是终止肽链合成并使肽链释放出核糖体。它们的性质见表 8-11。

表 8-11　蛋白质合成中的释放因子

释放因子	分子质量/Da	识别密码子
原核生物		
RF-1	44 000	UAA，UAG
RF-2	47 000	UAA，UGA
RF-3		本身无识别密码子性质
真核生物		
RF	150 000～250 000（2 个相同亚基）	UAA，UAG，UGA

RF-1 和 RF-2 各识别 2 个密码子（如表 8-11 所示），其中 UAA 相同；而 RF-3 本身不识别密码子，但可增加 RF-1 和 RF-2 活性。释放因子与核糖体和终止密码子能生成三元复合物。释放因子行使功能时可能需要 GTP，GTP 水解后使该因子从核糖体释放。

对真核生物释放因子研究较少，已知许多性质与原核相似，但唯一的一种释放因子能识别所有 3 种终止密码子。

二、蛋白质生物合成的机制

蛋白质生物合成可以分为三个步骤：①起始。核糖体亚基和起始 tRNA 在起始因子及其他因子参与下结合在 mRNA 编码区 5′端，保证起始 tRNA 识别起始密码子。结果生成起始复合物：核糖体·mRNA·起始 tRNA。②延伸。核糖体与 mRNA 相对移动，在延伸因子参与下由 tRNA 携带氨基酸（通过 aa-tRNA）进入核糖体，合成由 mRNA 序列编码的多肽链。所以延伸是一个循环过程（见图 8-3）。③终止。延伸至 mRNA 上出现终止密码子，释放因子进入核糖体，使新生肽链及核糖体从 mRNA 上释放出来。

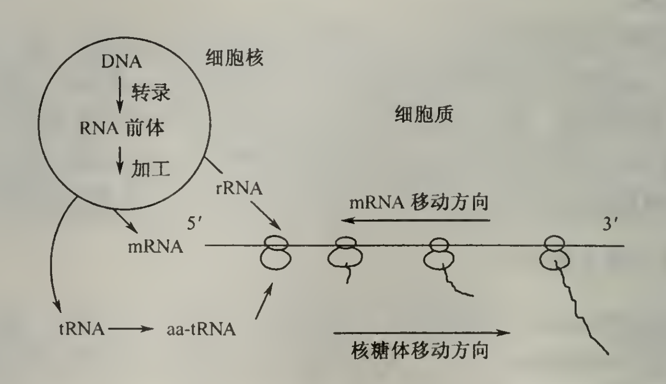


图 8-3　转录、加工与翻译的关系

1. 蛋白质合成的起始

无论原核生物还是真核生物蛋白质合成的起始都需要生成起始复合物：核糖体·mRNA·起始 tRNA。两者的区别在于核糖体及其亚基的大小、起始 tRNA、参与反应的起始因子和生成起始复合物的顺序不同，见表 8-12。

表 8-12 原核和真核生物蛋白质合成起始的区别

内容		原核生物	真核生物
核糖体	完整	70S	80S
	亚基	50S，30S	60S，40S
起始 tRNA		$fMet\text{-}tRNA_f^{Met}$	$Met\text{-}tRNA_i^{Met}$
起始因子		3 种	至少 7 种
起始复合物生成顺序		1.30S mRNA	1.40S Met $\text{-}tRNA_i^{Met}$
		2.30S mRNA·fMet $\text{-}tRNA_f^{Met}$	2.40S mRNA·Met-$tRNA_i^{Met}$
		3.70S mRNA·fMet $\text{-}tRNA_f^{Met}$	3.80S mRNA·Met-$tRNA_i^{Met}$

原核生物起始总反应式是：

$$30S+50S+mRNA+fMet\text{-}tRNA_f$$

$$\xrightarrow[GTP]{IF\text{-}1、IF\text{-}2、IF\text{-}3}[70S\cdot mRNA\cdot fMet\text{-}tRNA_f]+GDP+Pi$$

在一定浓度 Mg^{2+} 的生理条件下，核糖体亚基很易聚合，所以首先 70S 核糖体必须在 IF-1 和 IF-3 作用下解离。

$$70S=50S+30S\xrightarrow[IF\text{-}1]{IF\text{-}3}50S+30S\cdot IF\text{-}3$$

然后再形成 30S·IF-1、2、3 复合物。这 3 个因子有结合的协同作用，并且都结合在靠近 16S rRNA 的 3′端序列，位于核糖体的界面，从而妨碍了 50S 亚基的结合，形成 30S（小亚基）起始复合物：30S·IF-1、30S·IF-2、30S·IF-3＋mRNA＋$fMet\text{-}tRNA_f$＋GTP→［30S·IF-1、2·mRNA·$fMet\text{-}tRNA_f$·GTP］＋IF-3。共价交联实验表明，当形成 30S 复合物时，S7、S1、S11、S12、S18 和 S21 等蛋白质与核糖体识别 mRNA 起始位点有关。70S 起始复合物（简称起始复合物）的生成和 IF-2 的再循环通过以下步骤进行。50S 亚基的加入伴随着 GTP 被 IF-2（依赖于核糖体的 GTPase）水解，然后 IF-2（和 IF-1）及 GDP、Pi 从核糖体释放出来。

真核生物蛋白质合成的起始机制近年来也取得不少进展。真核生物起始的总过程见图 8-4。

1）生成 40S $Met\text{-}tRNA_i$·GTP 复合物。真核生物首先生成小亚基·起始 tRNA，然后再加入 mRNA，即 40S 亚基与 $Met\text{-}tRNA_i$ 结合在 mRNA 进入之前，这与原核生物有区别。$Met\text{-}tRNA_i$ 与 40S 结合是在 eIF-2 和 GTP 存在下进行的，eIF-2 与 $Met\text{-}tRNA_i$ 等量存在于 40S 亚基中。实验表明，在 $Met\text{-}tRNA_i$ 进入亚基之前，eIF-3 已存在于亚基中。

2）40S $Met\text{-}tRNA_i$·GPT 与 mRNA 结合。这是一个复杂的过程，至少与 3 个起始因子——eIF-3、4A、4B 有关，可能还与 eIF-1、4C 及 4E、4G 等起始因子

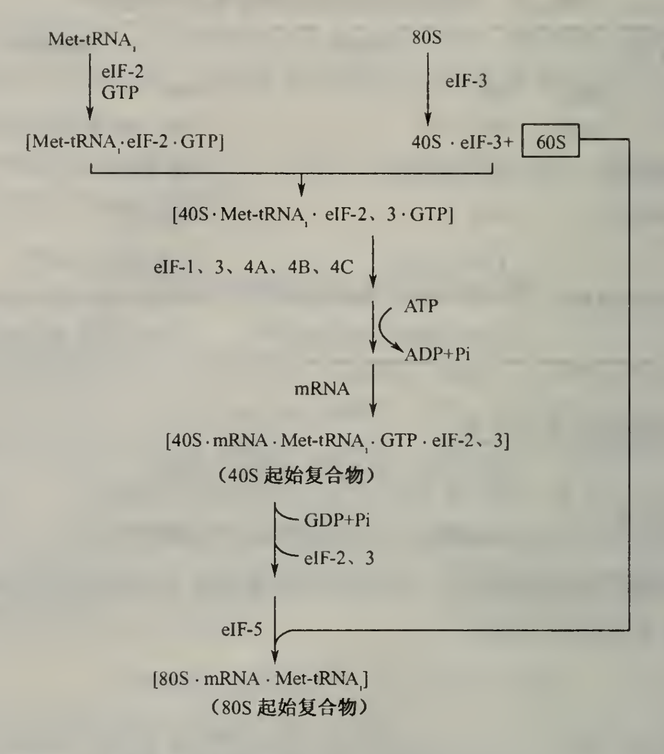


图 8-4　真核生物蛋白质合成的起始过程

有关。在 mRNA 与 40S 亚基结合时，mRNA 分子本身的结构特征起着重要作用。但在真核生物中，没有发现像原核生物中 16S rRNA3′ 端与 mRNA5′ 端非编码区形成多个碱基配对的情况。由于帽子类似物（m^7GMP）能在体外抑制带帽 mRNA的翻译，故真核 mRNA 的帽子结构在形成起始复合物中可能起促进作用。

3）60S 亚基加入到 40S mRNA·Met-$tRNA_i$ 复合物中。这个过程包括 eIF-2、eIF-3 的作用和 GTP 的水解，同时需要 eIF-5 参与 60S 亚基的结合。

2．肽链的延伸

肽链延伸过程由多个循环组成，每个循环包括：①氨酰 tRNA（aa-tRNA）与核糖体结合；②肽键形成；③移位。肽链延伸始于上述起始复合物的生成，除某些细节外，原核生物和真核生物的延伸过程大致相同。

1）aa-tRNA 与核糖体结合：在氨酰 tRNA 合成酶催化下，氨基酸被专一的 tRNA 所携带，生成 aa-tRNA；然后在 EF-Tu 和 GTP 作用下强烈促进 aa-tRNA 与核糖体结合。实验已证明，aa-tRNA 与核糖体的结合部位有 3 个，A 位、P 位及 R 位。aa-tRNA 首先结合于 R 位，然后经翻转至 A 位，见图 8-5。

2）肽键形成：aa-tRNA 结合于核糖体 A 位后，在肽基转移酶催化下，将 P 位 tRNA 上的肽基通过其羧基与 A 位 tRNA 上氨基酸的 α-氨基连接，形成肽键。真核生物中肽链形成与原核生物原则上相同，只是两类生物的肽基转移酶的活性中心可能存在结构差异。

3）移位：肽键形成后，核糖体处在前移位状态（pretranslocation state）；这时占领 P 部位的是去酰化 tRNA，A 位为肽基 tRNA，然后进行移位。移位包括 3 个相关步骤：核糖体与 mRNA 相对移动一个密码子的距离；肽基 tRNA从 A 位移至 P 位；去酰化 tRNA 离开 P 位。移位由 EF-G（原核生物）或 EF-2（真核生物）催化，随后 GTP 水解，导致 EF-G 或 EF-2 释放。

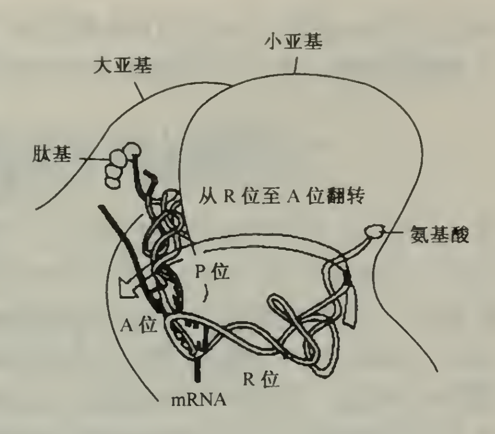


图 8-5　核糖体上三个 tRNA 结合部位——R、A 和 P 位

原核生物的肽链延伸循环见图 8-6。每次循环共需 3 个延伸因子（EF-Tu、Ts、G）和 2 分子 GTP，在 mRNA 密码子指导下，肽链延伸一个氨基酸残基。

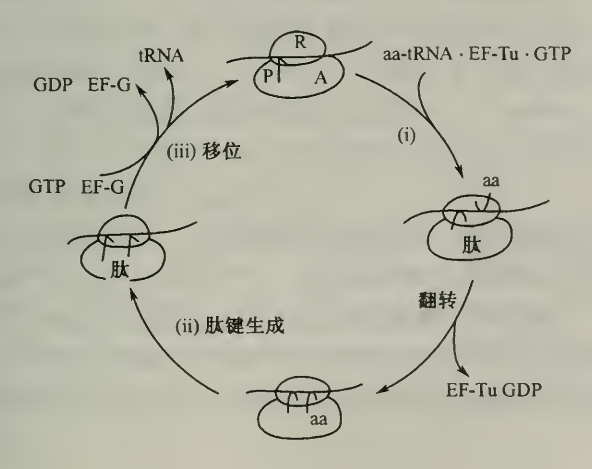


图 8-6　肽链延伸循环

3. 肽链的终止

当 mRNA 分子上终止密码子出现在核糖体的 A 位时，在一般情况下由于没有任何氨酰 tRNA 可以与终止密码子结合，不能把这些密码子翻译成氨基酸。释放因子（终止因子）在 GTP 存在下却能识别终止密码子；RF 与 A 位结合后，活化肽基转移酶，把 P 位上的肽基转移至水分子，即水解 P 位上 tRNA 与肽链之间的键；然后新生肽链与最后一个去酰化 tRNA 离开核糖体。RF 具有依赖核糖

体的 GTPase 活性，催化 GTP 水解使 RF 与核糖体解离。之后“空的”核糖体进入核糖体循环，图 8-7 是肽链终止过程示意图。

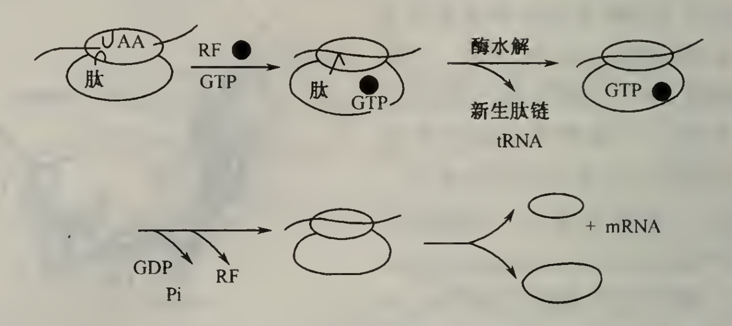


图 8-7　肽链合成终止过程

由 mRNA 翻译出来的多肽链是没有功能的，这叫新生蛋白质或蛋白质前体，需要经过加工改造才能成为有功能的蛋白质。

第九章　原核基因表达调控

为什么基因功能要调控？我们知道一个大肠杆菌基因组大约有 3×10^6bp，假如平均一个基因为 1000bp，330 个氨基酸为一个蛋白质，这样算起来，一个 *E.coli* 就有 3000 个基因，一个 *E.coli* 细胞总共含有 10^7 个蛋白质分子，如果每个基因都等同翻译的话，那么每个基因翻译的蛋白质分子应有 $10^7/3000\approx3000$ 个，但是对 *E.coli* 各种蛋白质的相对细胞含量的详细分析表明，细胞中有些蛋白质少至 10 个分子，而另一些蛋白质却多达 500 000 个分子，相差几乎是 5 万倍。这样的差距是如何产生的呢？这只能认为是有的基因不停地“工作”，而有些基因则缓慢地“工作”。的确，要使细胞的酶体系既协调又经济地发挥作用，每个基因产物的量最适宜，既不过多，也不过少，那么基因表达的调控就是必需的。

细菌中所利用的大多数基本调控机制遵守下列规则：一个体系在需要时被打开，不需要时被关闭。

这种“开”-“关”（on－off）活性是通过调节转录来建立的，也就是mRNA 的合成是可以被调节的。实际上，当说一个系统处于“off”状态时，也有基础水平的基因表达，常常是每世代每个细胞只合成 1～2 个 mRNA 分子，因而合成少量蛋白质分子，但为了方便，我们常常使用“off”这一术语，但必须明白所谓“关”实际的意思是很低（very low）。

细菌的某一代谢途径中，几个酶按顺序起作用，在这种情况下，往往是这一系列的酶要么都产生，要么都不产生，这种现象叫协同调节。实际上这是由于编码上述所有基因产物的多顺反子 mRNA 的合成被调节的结果。

酶活性通过生物合成途径中的产物或中间产物的浓度来共同调节，这个调节模式叫反馈抑制或终产物抑制。小的效应物分子也反复地用于激活或抑制一个特定的酶的活性。

第一节　乳糖代谢系统和操纵子模型

在乳糖的分解代谢系统中，代谢调节研究得最早最详细。F. Jacob 和 J. Monod 用遗传分析的方法建立了乳糖操纵子的负调控模型。这里主要介绍这一模型建立的研究过程及其实验依据。

大肠杆菌的乳糖转录子（lac scription），一般也称为乳糖操纵子（lac operon）包括三个结构基因：*Z*、*Y*、*A*。

P	O	Z	Y	A

转录方向

转录时，RNA 聚合酶首先与启动区（promoter）P 区结合，通过操纵区（operator）O 区向前转录。转录从 O 区的中间开始向 $Z\rightarrow Y\rightarrow A$ 方向进行，每次转录出来的一条 mRNA 上都有这三个基因的信息。转录的调控是在启动区（P 区）和操纵区（O 区）进行的。

三个结构基因各决定一种酶：Z 决定 β-半乳糖苷酶；Y 决定 β-半乳糖苷透性酶；A 决定 β-半乳糖苷转乙酰酶。

β-半乳糖苷酶是一种 β-半乳糖苷键的专一性酶，除能将乳糖水解成葡萄糖和半乳糖外，还能水解其他 β-半乳糖苷（如苯基半乳糖苷）。β-半乳糖苷透性酶的作用是使外界的 β-半乳糖苷（如乳糖）能透过大肠杆菌细胞壁和原生质膜进入细胞内。所以大肠杆菌如果以乳糖为碳源和能源的话，以上两种酶是必需的。β-半乳糖苷转乙酰酶的作用是把乙酰辅酶 A 上的乙酰基转到 β-半乳糖苷上，形成乙酰半乳糖。它在乳糖的利用中并非必需。

一、酶的诱导

在 19 世纪末，20 世纪初，科学家们已经知道微生物的酶的性质取决于微生物所生长的培养基，因此微生物能够被“训练”适应于各种不同的环境。

1900 年，F. Dienert 发现，将酵母放入以乳糖或半乳糖为唯一碳源的培养基时，酵母就能产生乳糖和半乳糖代谢所必需的酶，而再把这些酵母转到以葡萄糖为唯一碳源的培养基时，这些酶就消失了，因为在以葡萄糖为唯一碳源的培养基中这些酶对酵母菌的生长是不需要的了。在 20 世纪 30 年代，H. Karstrom 研究了细菌中碳水化合物代谢过程中几种酶的形成，并把它们分成两大类：一类是适应酶（adaptive enzyme），这种酶只在培养基中存在它们的底物时才能形成，另一类酶是组成酶（constitutive enzyme），无论培养基中是否有它们的底物，这种酶都能形成。适应现象是生物进化中所获得的，实际上，这是一种节约。

E. coli 的 β-半乳糖苷酶是一种适应酶。这种酶催化它的天然底物乳糖以及其他 β-半乳糖苷化合物的水解。当 *E. coli* 以乳糖为唯一碳源时，这种酶就产生，而当培养基中不存在乳糖时，这种酶就消失。

后来发现有些物质虽不是酶的底物，但可高效地诱导酶的产生，有些 β-半乳糖苷是酶的底物但不是诱导物，不诱导酶的产生。所以后来把适应酶改为诱导酶，并把其存在而能引起细胞产生一种酶的这些化合物定名为诱导物。

1946 年，Jacques Monod 开始了关于 *E.coli* β-半乳糖苷酶的适应形成的研

究，15 年后，这个研究解决了细菌基因功能的调控问题。在这个研究的第一阶段，Monod 和他的同事们就将他们的工作集中到问题的焦点，并确定了酶适应过程的真正本质。1953 年在他们发表的论文中，将适应酶改为诱导酶。

在对 *E.coli* 形成 β-半乳糖苷酶的诱导物的性质的研究中发现，普通的 β-半乳糖苷的不可代谢的含硫类似物，如甲基硫代半乳糖苷（Thiomethyl galactoside, TMG）和异丙基硫代半乳糖苷（isopropythiogalactoside）是高效的诱导物。

图 9-1 是 *E.coli* β-半乳糖苷酶的两个不可代谢的即不消耗的诱导物异丙硫基半乳糖苷（IPTG），以及酶分析中所用的产颜色的酶底物 O-硝基苯半乳糖苷（ONPG）的结构。

这种高效诱导物可以用来研究在不以乳糖为碳源和能源的培养基中，在不消耗诱导物的情况下的诱导过程的真正动力学。在这种情况下细菌的生长速度与 β-半乳糖苷酶的浓度无关，并且诱导物的浓度保持不变。通过这个研究发现，在向生长培养基中加入饱和量的诱导物后的几分钟内，在不消耗诱导物的情况下，细胞按一个恒定的最大速度 *P* 合成 β-半乳糖苷酶。从培养基中把诱导物一除去，*P* 立即就下降到最大诱导物浓度存在时所达到的速度的千分之一以下。因此显然在细菌生长期间，这种新合成的原生质中有恒定的一部分 *P* 是 β-半乳糖苷酶蛋白质，该 *P* 值取决于生长培养基中的诱导物浓度（图 9-2）：正如图 9-2 所示，在诱导物浓度最大时，*P* 值是 0.066，这就是说，被最大限度诱导的细胞里，所制造的全部蛋白质中有 6.6% 是 β-半乳糖苷酶。

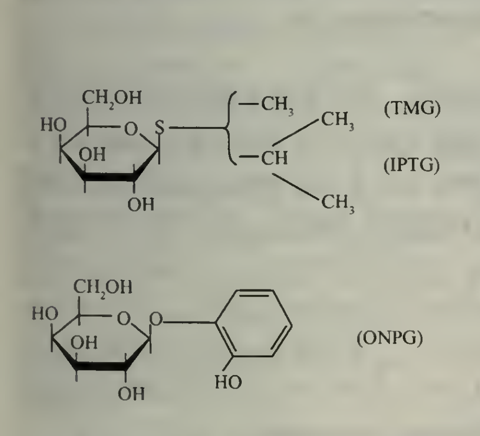


图 9-1　β-半乳糖苷酶的高效诱导物

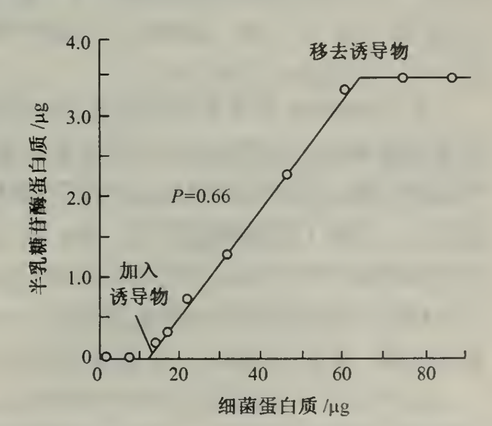


图 9-2　在不消耗的情况下，*E.coli* 的 β-半乳糖苷酶的诱导动力学

为了证明诱导物的作用是诱导新合成酶而不是将已存在于细胞中的酶前体物转化成有活性的酶，设计了同位素示踪实验。把 *E.coli* 细胞放在加有放射性 ^{35}S 但没有任何半乳糖苷诱导物的培养基中繁殖几代，然后再将这些带有放射活性的细菌转移到不含 ^{35}S 的无放射性的培养基中，随着培养基中诱导物的加入，β-半

乳糖苷酶便开始合成。分离纯化β-半乳糖苷酶，发现这种酶无^{35}S标记。说明酶的合成不是由前体转化而来，而是加入诱导物后新合成的。

二、结构基因和调节基因的突变

Monod 在 *E.coli* 中开始关于乳糖发酵作用的生理学研究后不久，J.Lederberg 便开始探讨它的遗传基础。1948 年 Lederberg 分离了一系列 *E.coli* 的乳糖发酵负（*lac*）突变型，这些突变型不能发酵乳糖。对这些突变型进一步研究证明，其中一些突变型的 lac^- 特征是由于不能产生β-半乳糖苷酶。在连续的几年内，越来越多的 lac^- 突变型分离出来了，了解了它们的特性，并且，通过细菌结合转移和转导实验将它们的突变位点定位在细菌遗传图上。其中一些 lac^- 突变型的生理基础显得很复杂，而另一些则比较简单。最简单的两种类型，一种是 $lac\ Z^-\ Y^+\ A^+$ 是β-半乳糖苷酶基因突变；另一种是 $lac\ Z^+\ Y^-\ A^+$ 是透性酶基因突变（$lac\ Z^+\ Y^-$ 突变型称为隐秘突变型，因为它的水解乳糖的能力在完整细胞内不显示出来，只有在细胞提取物中才表现出来）。鉴定出所有这些突变后，通过研究证明这三个基因在染色体上是连锁的，见图 9-3。

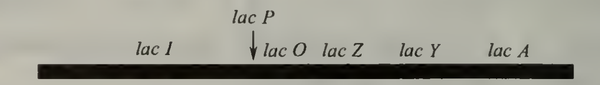


图 9-3　*E.coli* 中乳糖发酵基因的次序

在 Lederberg 的突变型中还有另一类突变型，其半乳糖苷酶的形成不依赖于诱导物的存在，这种组成型突变体在没有乳糖或别的半乳糖苷存在时也能合成β-半乳糖苷酶。但野生型则必须在存在诱导物时才能合成这种酶。组成型突变分两大类；一类为 I 型组成型突变，另一类为 O 型组成型突变。为了和野生型相区别，两类突变型分别以下列符号标记：

I 型：野生型为 I^+，突变型为 I^-；

O 型：野生型为 O^+，突变型为 O^c。

这样的突变可以说是多效性的突变，$I^+ \rightarrow I^-$ 或 $O^+ \rightarrow O^c$ 后，后边的 *ZYA* 结构基因均表现为组成型。*I* 基因叫调节基因（regulatory gene)。

三、调节基因

在没有任何外源诱导物时，$lac\ I^-$ 组成型突变型是如何合成它们的 *lac* 基因的酶呢？一种最方便的解释是从 $lac\ I^+$ 到 $lac\ I^-$ 的突变导致了一种内源诱导物的形成，因而这种内源诱导物使细胞能在没有外加半乳糖苷时合成它们的酶。但事

实上，这种内源诱导物是不存在的，因为在组成型突变中从来就没有找到过这样的内源诱导物。这样看来，研究 I^- 的作用就十分重要了，如何证明 I 基因的调节作用呢？我们可以设法组成一个部分二倍体细胞，考查 I^+、I^- 的行为，就可以知道调节基因在细胞中的作用。

20 世纪 50 年代后期，Wollman、Jacob 和 Hayes 的工作，已揭示了在细菌结合过程中，一个雄的 Hfr 供体细菌的 DNA 转入雌性的受体细菌（F^-）后，可形成暂时的部分合子，根据这个道理，A.B.Pardee，Jacob 和 Monod 设计了一个实验，来探讨 I 基因在细胞中的作用。他们将 Hfr lac I^+ Z^+ $str^S T6^S$ 作供体菌，F^- $lacI^- Z^- str^r T6^r$ 作受体菌，分别在无任何诱导物的培养基上培养后，再混合在一起做杂交实验，图 9-4 是实验结果。

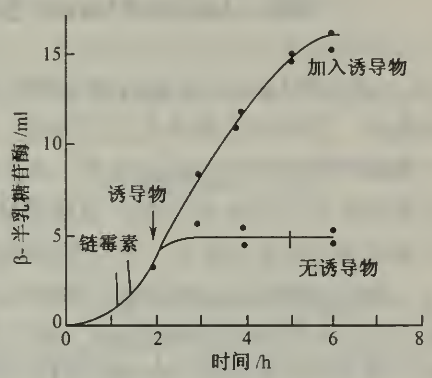


图 9-4 由 *lac* $I^+ Z^+ Str^s$，$Tsx^s Hfr$ 供体 *E.coli* 与 $lacI^- Z^-$，str^r，Tsx^r，F^- 受体结合转移后 β-半乳糖苷酶合成情况

由实验结果可以看出，当供体的 *lac* I^+ Z^+ 基因进入受体之前，由于供体没有接触诱导物，受体是 *lac* Z^-，所以二者都不合成 β-半乳糖苷酶，结合进行 1h 后加入 T6、str 杀死供体菌，从 1h 到 2h 之间，β-半乳糖苷酶活性逐渐上升，但 2h 后，不加诱导物则酶的合成就停止，在这个实验的对照组，结合 2h 加入诱导物，则酶的合成继续进行。对照结果说明，部分合子在主要实验组中最后停止合成酶不是因为它没有合成酶的能力，而是成了诱导型了。

为什么结合进行到 1h 时，*lac* 的酶组成型合成，而到 2h 时却又诱导型合成呢？可以认为结合需要 1h，*lac* I^+ Z^+ 才转移到受体菌中，由于受体菌中原来并无 *lac* I 产物，因此供体菌转移进来的 *lac* Z 基因可以组成型表达，但在 *lac* Z 基因表达的同时，*lac* I 基因也在表达。随着时间的延长（2h 后），I^+ 产物积累达到足以阻止 *lac* Z 基因的转录时，部分合子就从组成型变为诱导型了。从对照组的结果还可以看出诱导物是使 *lac* I 产物失活，因而加入诱导物后，*lac* Z 基因继续表达。

在提出有 *lac* I 抑制物这一假设以后的许多年内，对于进一步确定它的性质取得的进展很小，后来选到了 *lac* I 的温度敏感突变株和无义突变株，这就说明 *lac* I 基因产物是一种蛋白质。后来，还发现了 *lac* I 的超阻遏突变株（*lac* I^S），它的酶的合成诱导需要的诱导物的浓度比野生型要高得多（或者完全不被诱导，表现为 lac^-）。

由上述实验及结果分析，可以建立下述模型：①I^+ 合成一种特异性的阻遏物，在没有诱导物时，它阻止 Z 基因的表达。②诱导物作为阻遏物的拮抗物起

作用，因而加入诱导物以后阻遏物失活，Z 基因得以表达。③I^- 不产生阻遏物，因而无需诱导物 Z 基因即可表达，阻遏物起负调控作用。

四、Jacob-Monod 的负控制模型及实验依据

Jacob 和 Monod 在 1961 年发现了负控制系统，在生物学发展史上有划时代的意义。发表时生物界反应之强烈，较之 Watson 和 Crick 在 1953 年发表 DNA 双螺旋模型时有过之而无不及。他们这篇经典论文中还预言了 mRNA 的存在，从而导致了 mRNA 的证实，使氨基酸密码的实验研究得以开始，结果使整个分子遗传学体系建立起来，所以它的历史意义是非常重大的。

他们的研究目的就是要了解基因功能的调控，这是整个生物界最根本性的问题。但几十年来没有丝毫突破，他们看出了细菌中的适应酶现象是个基因调控问题，而且可以用实验方法进行分析，因此他们选它作为突破口，终于通过大量实验及分析，建立了乳糖操纵子的负控制模型（图 9-5）。

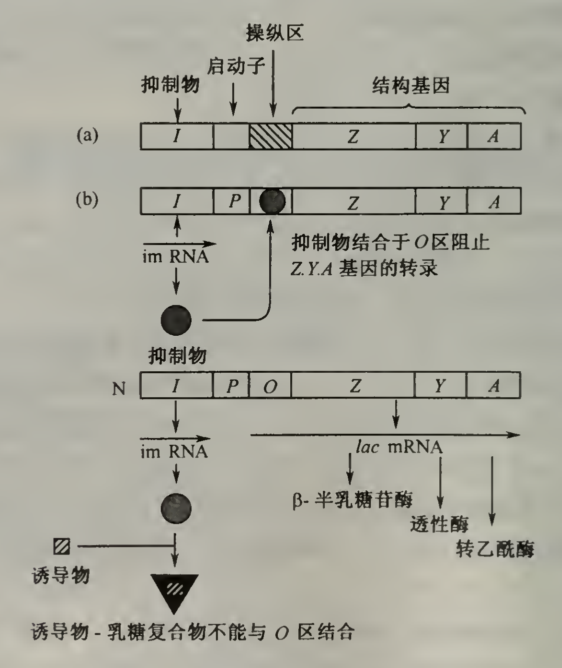


图 9-5　乳糖操纵子负调控模型

图 9-5 说明①I 基因产生一种抑制物，这种抑制物是一种可以变构的蛋白质，它有两个结合位点，当没有诱导物时，可与 DNA 上 O 区结合，因而关闭 Z、Y、A 基因的转录。②当有诱导物时，诱导物与抑制物结合，改变了抑制物的构型，使它不能与 O 区结合，因而打开 Z、Y、A 基因的转录。

当 I 基因突变时，因不能合成抑制物而使 Z、Y、A 基因能转录，$O^{+} \rightarrow O^{c}$ 时，是由于 O 区结构发生变化而使抑制物不能与之结合，也使得 Z、Y、A 基因能转录。

这个模型非常合理地解释了 *E.coli* 乳糖发酵中的酶的诱导现象的本质。

Jacob 和 Monod 分离了一些组成型突变株，用细菌结合转移方法制成部分合子，观察 β-半乳糖苷酶的合成情况，为他们的负控制模型提出了令人信服的实验依据。

组成型突变：①$\begin{cases} I^{-} \\ O^{c} \end{cases}$ β-半乳糖苷酶组成型合成　②$Z^{-}\begin{cases} \text{β-半乳糖苷酶不合成} \\ \text{CRM（免疫交叉反应物）产生} \end{cases}$

部分二倍体：$\frac{O^{+}Z^{+}}{O^{c}Z^{-}}$（$O^{+}Z^{+}$ 为 β-半乳糖苷酶诱导型，$O^{c}Z^{-}$ 为 CRM 组成型）；$\frac{O^{+}Z^{-}}{O^{c}Z^{+}}$（$O^{+}Z^{-}$ 为 β-半乳糖苷酶组成型，$O^{c}Z^{+}$ 为 CRM 诱导型）；$\frac{I^{+}Z^{+}}{I^{-}Z^{-}}$（$I^{+}Z^{+}$ 为 β-半乳糖苷酶诱导型，$I^{-}Z^{-}$ 为 CRM 诱导型）。以上部分合子的不同行为充分说明了 O 型组成型突变是顺位显性反位隐性，而 I^{+} 基因对于 I^{-} 来说是顺位反位都是显性。

以上结果分析完全符合负调控模型，负调控就是由于某基因或基因产物的存在导致转录不进行，酶的合成受抑制，就叫负调控。

五、*I* 基因产物及功能

前面部分二倍体实验已证明 I 基因能产生一种在细胞质中扩散的物质，后来分离到了 I 基因的 ts 和 am 突变型，这就表明了 I 基因产物是一种蛋白质，后来又分离到了 I^{S} 突变型，表型为 Lac^{-}，同时另一些 *lac I* 基因突变型，它们的抑制物在有诱导物时对热的稳定性升高，这就预示了 I 基因产物可能是与诱导物相互作用的。

1967 年 W.Gilbert 和 B.Miller.Hill 终于鉴定出 *lac I* 基因的产物。他们检验了 *E.coli* 的浓缩提取物中能够与放射性标记的异丙基硫代半乳糖苷（IPTG）结合的一种蛋白质。结果发现这种蛋白质很少，相当于每个细胞只结合约 20～40 个 IPTG 分子，这种蛋白质能提纯到比它的初始浓度高 100 倍左右，同时通过测定这种蛋白质在离心场内的沉降速度，发现这种蛋白质分子大约含有 1000 个氨基酸，并且能与 IPTG 结合。因此在每个 *E.coli* 细胞中大约有 5～10 个这种蛋白质分子。在 *lac* I^{-} 组成突变型中不能检出这种蛋白质，说明这种结合 IPTG 的蛋白质确实是 *lac* I^{+} 基因所产生的抑制物。将这种能与 IPTG 结合的 I^{+} 基因产

物在离心场中测沉降常数，测得其为7.8S的物质，用核酸酶与之作用数小时后，其活性不失去，而用链霉蛋白酶50℃作用可失去活性，这就确定地说明 *I* 基因产物是蛋白质。

为了能够对Jacob和Monod的关于 *lac O* 基因是抑制物的作用目标进行鉴定实验。W.Gilbert 和 B.Miller-Hill 从^{35}S标记了的细菌中提取 *lac I* 基因的产物——抑制物，并且在离体的反应混合液中有或没有IPTG的2种情况下，把放射性标记的抑制物蛋白质与不同的DNA分子相互作用，然后，将这些混合物进行蔗糖密度梯度沉降以确定是否有任何放射性抑制物蛋白质（这种蛋白质在游离状态下，沉降速度仅为实验中用的DNA分子的五分之一）已结合到DNA分子上以同样速度与DNA一起沉降。这个实验结果如下：①抑制物不与失去了 *lac* 操纵子全部基因的DNA分子结合，②抑制物结合于带有 *lac* 操纵子基因的DNA分子上。但必须先除去IPTG诱导物，有IPTG便会阻碍抑制物与DNA结合。③如果操纵子是操纵区组成型 *lac* O^c 突变体，则抑制物不与此DNA分子结合。因此，这些结果直接地肯定了最初关于抑制物作用机制的假设。

由于后来选到了 *lac I* 的另一种突变株，能产生大量的 *I* 抑制物，用这种突变株可以制备大量的 *lac I* 抑制物，测定其分子质量为150 000Da，有四个亚基，每个亚基37 000Da，只有四聚体才有功能，每个亚基由347个氨基酸组成，每个亚基都有与诱导物和 *O* 区的结合位点。

用胰蛋白酶、胰凝乳酶处理此亚基，它丧失了与 *O* 区结合的能力而不失去与诱导物结合的能力。分析此亚基的结构，发现1～59，328～347氨基酸失去了，说明1～59，328～347是与 *O* 区结合的部位。

六、操纵区和启动区

怎样知道结构基因 *Y* 和 *Z* 都控制着乳糖发酵特性呢？这是因为不但 *Y* 突变后，细菌由乳糖发酵变为不发酵乳糖，而且发现 *Y* 不变而 *Z* 突变后，细菌同样由发酵乳糖变为不发酵乳糖。

i 基因突变后，细菌从诱导型变为组成型。在 *i* 基因不变时，也曾发现细菌从诱导型变为组成型。这说明有另一基因参与细胞中诱导酶合成的调节控制。这便是操纵区，它的组成型写作 O^c，通过杂交，可以看到 *O* 和 *i* 在染色体上有不同的位置，而 *O* 和 *Z* 则是邻接的。

转录以RNA聚合酶和DNA特定位置结合开始，在乳糖操纵子中，这一结合位置在哪里？

如果RNA聚合酶和阻遏物的结合位点是一个，那么这一部位的缺失使聚合酶不能发生作用，因此缺失应是导致超阻遏型，阻遏物结合部位的缺失则导致组成型。缺失突变型的研究结果说明确实存在这两种突变型。从而在阻遏物结合部

位（操纵区 *O*）以外，还鉴定出了 RNA 聚合酶结合部位（启动区 *P*）。

1975 年以前，启动子的碱基顺序的性质还不完全清楚，而现在，许多启动子的顺序已经测定。通过比较，发现在 *P* 区有些共同的序列，就是在转录一章中所讲的 −10 区（pribnow box）和 −35 区（−35 sequence），有些启动子上还有辅助因子结合位点，如 cAMP-CAP 结合位点，乳糖操纵子的 *P* 区就包括以上三个结合部位。其中 cAMP-CAP 与 DNA 结合起正调控作用。

七、正控制系统

cAMP 是 Sutherland 等在 20 世纪 50 年代发现的，多年来不知道它的生物学意义，未受到重视，直到 1965 年发现脊椎动物的几种肽激素作用于靶组织时，首先使靶组织的 cAMP 的含量增高，这才引起人们的重视，现在已经清楚 cAMP 在激素调节中起第二信使作用，Sutherland 由于这个发现而获得诺贝尔奖。

1965 年 Sutherland 发现了大肠杆菌的培养液中葡萄糖的含量总是与 cAMP 的含量成反比。当时还没有立即想到 cAMP 与“分解代谢物抑制”的关系，直到 1968 年，发现培养液中加入 cAMP 后可以克服葡萄糖对 β-半乳糖苷酶的抑制作用，也就是说在培养液中加入 cAMP 可以增加 β-半乳糖苷酶的产量，而把生物学中表面上看来无关的两个重大问题联系起来，从而导致正控制系统的发现。

在正调控系统中，在有诱导物（抑制解除物）存在的条件下，细胞内 3′、5′环式 AMP（简写 cAMP）的合成量就多，cAMP 量多则 *lac* 操纵子转录就愈盛。cAMP 的调控作用是通过“cAMP 受体蛋白”CAP（CAP 或称 CRP 是正控制中的一个必需的蛋白质因子）和 cAMP 结合成复合物（cAMP-CAP 复合物），此复合物和 *lac* 操纵子 *P* 区的 CAP 位点结合，转录才能开始。cAMP-CAP 与 P 区结合，改变 DNA 的次级结构，使 RNA 聚合酶能与 −10 区形成 open-promoter complex，这一机制称为正控制。

正控制系统的要点是 cAMP-CAP 对转录的诱导作用（induction），未和 cAMP 结合的 CAP 是“无活性的”诱导物，它不能和 CAP 位点结合，已和 cAMP 结合的 CAP 是有活性的诱导物（inducer）；cAMP 称为“辅诱导物”（co-inducer）。

细胞内 cAMP 的含量受许多条件的控制，cAMP 是由 ATP 经腺苷酸环化酶（adenylcyclase）环化而成：ATP $\xrightarrow{\text{腺苷酸环化酶}}$ cAMP + 焦磷酸

各种因素都可以影响腺嘌呤核苷酸环化酶的活性，因而影响 cAMP 的含量，从而影响转录。例如，培养液中如果存在葡萄糖就可以影响环化酶的活力，使细胞内 cAMP 含量降低，从而抑制 *lac* 操纵子的转录。正控制系统中葡萄糖这样的抑制物称辅抑制物，这是为了和负调控系统中的蛋白质抑制物相区别。正控制和

负控制的比较见表 9-1。

表 9-1 正控制和负控制的比较

	正控制体系	负控制体系
主要作用物（蛋白质性质）	cAMP-CAP 复合物	抑制物蛋白
主要作用物的作用	诱导	抑制
作用位置	*P* 区	*O* 区
起相反作用的辅助因子 （非蛋白质性质）	辅抑制物 （葡萄糖）	抑制解除物 （半乳糖）

并不是葡萄糖本身直接起抑制作用，而是它的分解代谢物起作用。在大肠杆菌中，葡萄糖可以抑制 β-半乳糖苷酶的产生。这个现象很早就知道，传统中称为“分解代谢物抑制”，很多微生物都专一地利用一种糖，但 *E.coli* 可以利用葡萄糖也可以利用乳糖，在培养基上同时存在葡萄糖和乳糖时，*E.coil* 只利用葡萄糖，而不利用乳糖。

这可能是长期自然选择的结果，因为自然界中葡萄糖比其他糖存在广泛。另外，这显然也有适应意义。如果环境中有葡萄糖，细菌就可以利用葡萄糖作能源和碳源。因此细菌不产生 β-半乳糖苷酶是合理的节约，这也是通过自然选择而形成的极其合理和巧妙的调控机制。

八、*P-O* 区的结构

lac 操纵子早已有人分离得到，其 *P-O* 区的核苷酸顺序已全部分析清楚，*lac* 操纵子的 *P-O* 区结构如图 9-6 所示。

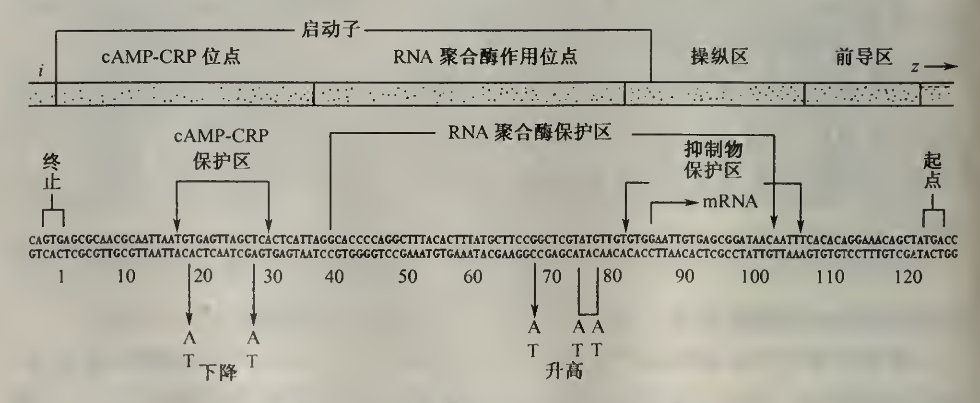


图 9-6 *E.coli* 乳糖操纵子调控区的碱基顺序

P 区是从 *i* 基因结束到 mRNA 转录开始点为止，而 *O* 区的范围是抑制物结

合区。从 *i* 基因结束到 *Z* 基因起点之间共 122 个碱基对，mRNA 是从第 85 对碱基开始的，所以 *P* 区共 84 个碱基对，抑制物结合区从 78～112，共 35 对碱基，和 *P* 区有 7 对重叠，抑制物保护区之后还隔 10 对才是 *Z* 基因的始点，现在一般把 mRNA 开始到结构基因起始密码子之间这一段称为前导区（leader）。

P-O 区碱基对的命名顺序是从 mRNA 起始的第一个核苷酸为 +1 与转录方向一致的为正，与转录方向相反的为负。所以 *P* 区从 -84～-1，抑制物保护区是 -7～28。

O 区（-7～28）的碱基顺序有对称性，其对称轴在 +11 碱基对（图 9-7）。*P* 区中的 cAMP-CAP 复合物结合区也是一个对称区（67～52），其对称轴在 60～59之间，此对称序列如图 9-7：对称区在蛋白质与 DNA 的识别结合中有重要意义。

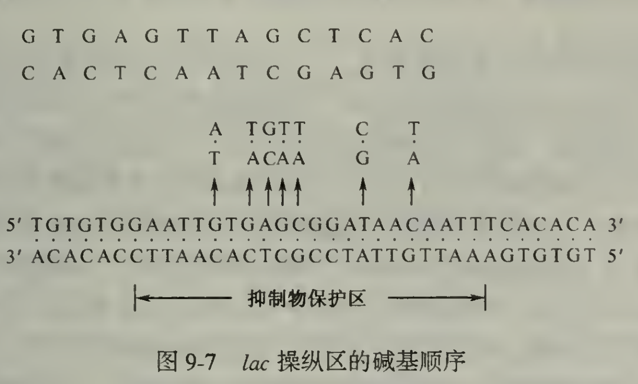


图 9-7 *lac* 操纵区的碱基顺序

第二节 半乳糖操纵子

大肠杆菌半乳糖操纵子（gal operon）在大肠杆菌遗传图上位于 17min 处，包括三个结构基因：*galE*（异构酶 UDP-galactose-4-epimerase），*galT*（半乳糖 1-磷酸尿嘧啶核苷转移酶，galactose transferase），*galK*（半乳糖激酶，galactose kinase）。这三个酶的作用是使半乳糖变成 1-磷酸葡萄糖。半乳糖操纵子的调节基因是 *galR*，位于遗传图上 55min 处（图 9-8）。

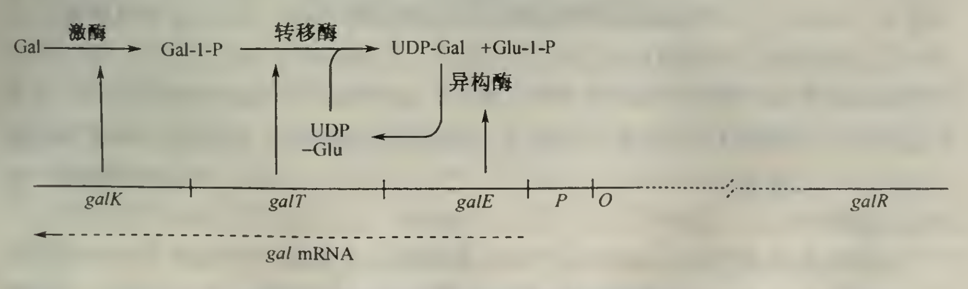


图 9-8 半乳糖操纵子，调节基因距结构基因很远

gal operon 与 *lac* operon 所不同的是 *galR* 与 *gal ETK* 基因族离得很远，*galR* 基因产生阻遏物蛋白，而 *galR* 基因与 *gal* 的结构基因和 *galO* 都不相邻，但 *galR* 产物和 *galO* 的作用与 *lac* 操纵子中的 *laci* 产物和 *lacO* 作用相同，因为 $galR^-$ 和 $galO^c$ 突变表型分别是可以组成型地合成 *E*、*T*、*K* 基因产物。*gal* 操纵子的诱导物常常是半乳糖。半乳糖操纵子还有两点与乳糖操纵子不同：①半乳糖操纵子有两个启动子，mRNA 从两个不同的起始点开始转录；②半乳糖操纵子有两个 *O* 区，一个 *O* 区不在 *P* 区与结构基因之间，而是在 *P* 区上游，在 −67～−53 的地方；另一个 *O* 区在结构基因内部，在 *galE* 基因内。现已知的所有操纵子中仅此一例。

一、cAMP-CAP 对两个 *gal* 启动子的不同作用

半乳糖的利用效率比葡萄糖要低。人们料想可能是葡萄糖存在时半乳糖操纵子不被诱导，但实际上在葡萄糖存在时，半乳糖操纵子是可以被诱导的。分离到一些突变株，奇怪的是，这些突变株可被分为两类，一类突变是能高水平地合成半乳糖代谢的酶，但当葡萄糖存在时不合成；另一类突变是当葡萄糖不存在时不能合成半乳糖代谢的酶，这些现象只能用半乳糖操纵子上存在两个启动子来解释。对半乳糖操纵子的 *P-O* 区进行碱基顺序分析，发现有两个启动子，相距 5 个 bp，半乳糖操纵子可以从两个启动子启动转录，有两个转录起始点 S_1 和 S_2，两个不同的启动子是 *PG*1 和 *PG*2。cAMP-CAP 对从 S_1 和 S_2 的起始转录的作用不同。

1）从 S_1 起始的转录只有当不存在葡萄糖时才能进行，当腺苷酸环化酶突变（cya^-）或 cAMP 受体蛋白（crp^-）突变时，半乳糖操纵子不能从 S_1 转录，在体外系统中加入 cAMP-CAP 时，才能从 S_1 能转录。

2）从 S_2 的转录是在葡萄糖存在的情况下起始的，从 S_2 的转录不需要 cAMP-CAP，事实上，cAMP-CAP 抑制从 S_2 的转录。

在半乳糖操纵子内的 cAMP-CAP 所起的调节作用比在乳糖操纵子中更为复杂，大肠杆菌的 cya^- 或 crp^- 突变型不能利用乳糖，但可以利用半乳糖作唯一碳源。那么是不是 cAMP-CAP 对半乳糖操纵子完全没有作用呢? 用体外转录的手段证明 cAMP-CAP 是能够刺激半乳糖操纵子增强转录的，用含有半乳糖操纵子调节区的限制酶片段作模板进行体外转录时，发现转录可以从 S_1 或 S_2 两个不同的起始位点开始，当有 cAMP-CAP 时转录从 S_1 开始，当无 cAMP-CAP 时，转录从 S_2 开始，说明 cAMP-CAP 抑制从 S_2 的转录而刺激从 S_1 的转录。顺序分析表明有两个相互重叠的 Pribnow box 顺序，它们分别对应于两个不同的起始位点 S_1 和 S_2（图 9-9）。

活体实验结果支持 cAMP-CAP 对半乳糖操纵子中的两个启动子的作用不同这一观点。发现有一个半乳糖启动子突变型不能利用半乳糖，但如果将这一突变

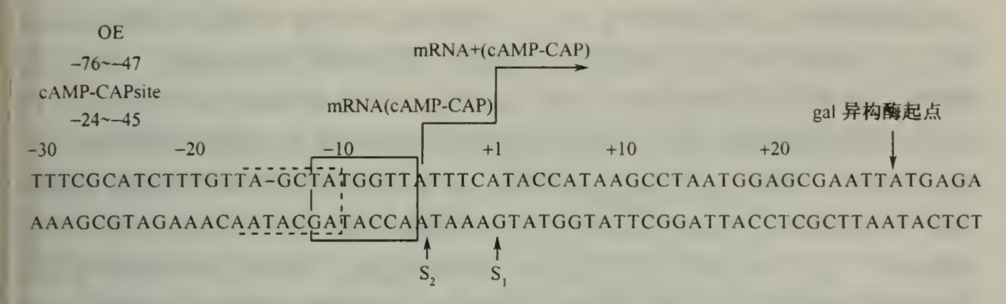


图 9-9　半乳糖操纵子转录起始点上游的碱基顺序，
有两个 Pribnow box 和两个转录起始位点 S_1 和 S_2

菌株再行突变为 *cya*⁻或 *crp*⁻时则又能够利用半乳糖了。原来的突变型在体外条件下不能从 S_1 开始转录，但能从 S_2 开始转录。说明这一突变型之所以不能利用半乳糖是由于突变损坏了在 S_1 处开始的转录（可能是 $PG1^-$），而细胞内又存在 cAMP-CAP，对 S_2 开始的转录有抑制作用。所以这种突变在活体内 gal 操纵子既不能从 S_1 转录也不能从 S_2 转录。因而不能利用半乳糖。这一突变再行突变后，不论是 *cya*⁻还是 *crp*⁻突变，都可以恢复对半乳糖的利用。

对这一突变型启动子核苷酸顺序分析结果支持上述观点，这一突变发生在 -11碱基对，使原来的 A/T 变为 T/A。这对碱基恰好在 *PG*1 的 Pribnow box 的第二对碱基上，已知许多启动子中这对碱基保守性极强，碱基改变必然会带来功能改变，因此不能从 S_1 处转录。此外，这对碱基也正好是 *PG*2 的 Pribnow box 中的第 7 个碱基，在其他启动子上，这对碱基变化程度也比较大，同样，这里也没有影响 *PG*2 的功能。

cAMP-CAP 是如何促进从 S_1 的转录并同时抑制 S_2 的转录呢? 这一点还不十分清楚，一般认为，促进 S_1 的方式和乳糖操纵子中的方式相似（即促进形成 open-promoter complex），并且在 *PG*1 启动子上形成的这个复合物在空间位置上干扰了从 S_2 的起始（因为该复合物和 S_2 的 Pribnow box 有两个碱基重叠）。

二、双启动子的生理功能

对于具有单启动子的基因来讲，启动子的生理功能容易理解。可是某些基因具有不止一个启动子，例如半乳糖操纵子上有两个启动子，这些启动子究竟具有什么生理功能呢?

半乳糖对细菌有双重作用，半乳糖一方面可以作为碳源，另一方面它又是细胞壁的成分。细胞壁物质的直接前体是 UDP-半乳糖，当细菌在不含半乳糖的培养基上生长时，可以通过差向异构酶将 UDP-葡萄糖转化为 UDP-半乳糖，从而

合成细胞壁物质。因此细胞中必须有一定量的差向异构酶才能维持细菌的生长和繁殖。这也说明为什么在不加诱导物时也有一定量的酶形成的必要性。事实上乳糖操纵子在诱导和不诱导时表达相差1000倍，而半乳糖操纵子则只相差10～15倍。当细菌以葡萄糖作为唯一碳源和能源时，细胞内缺少cAMP，于是*PG*1不能启动转录，但有利于*PG*2启动转录，因为*PG*2的启动不但不需要cAMP-CAP，而且还避免受cAMP-CAP对*PG*2的抑制作用。由于*PG*2的启动，于是细胞内有一定水平的有关半乳糖利用的三种酶的生成，其中的差向异构酶可以催化UDP-葡萄糖转化为UDP-半乳糖。当细胞中以半乳糖作碳源时，细胞中cAMP浓度升高，抑制*PG*2启动子，促进*PG*1启动子的启动，从而保证细菌能在两种不同的环境中都可以形成细胞壁物质的前体——UDP-半乳糖。

三、双操纵区

半乳糖操纵子与乳糖操纵子不同，前者的operator不在*P*区与结构基因之间，而是在*P*区的上游。考察半乳糖操纵子中−70～−50之间包含的一段回文对称顺序和乳糖操纵子上相对应位置的CAP位点非常相似，但又不是真正的CAP位点。因为半乳糖操纵子上−66处发生G/C→A/T突变后，操纵子上几个基因不经诱导就可以高水平表达，说明这一位点是操纵区而不是CAP位点。而且这一突变型体外转录仍需要cAMP和CAP蛋白，这就使我们确信半乳糖操纵子内这一序列虽然位置和结构都和乳糖操纵子的CAP位点相似但并不是CAP位置，实际上CAP位置在−24～−50之间，在操纵基因下游，近年来发现在半乳糖操纵子的CAP位点下游还有另一个操纵区，在半乳糖操纵子的*galE*基因内找到一株O^c突变，顺序分析表明*galE*内+46～+60的一段顺序与−53～−67有同源关系，体外结合实验表明可以结合抑制物。上游的*O*区称为外操纵区（*OE*），*galE*内的*O*区称为内操纵区（*OI*），两个操纵区顺序如图9-10所示，目前对于两个*O*区的功能还没有很完美的解释。

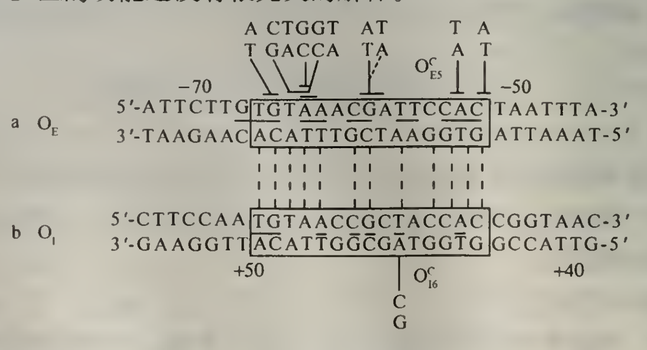


图9-10 *gal*操纵子的两个操纵区顺序

第三节　色氨酸操纵子

色氨酸操纵子（trp operon）是色氨酸合成的一个操纵子，操纵子的调控仅仅根据生长培养基中有无色氨酸，当培养基中有色氨酸时，不需要色氨酸操纵子处于活性状态，而当培养基中无色氨酸时，色氨酸操纵子被打开，这里就存在一个调控系统。

色氨酸的合成分五步完成。每一步需要一种酶，这五种酶的基因彼此相邻，是连锁在一起的，这五个基因被转录在一条多顺反子 mRNA 分子上，五个基因分别是 *trpE*、*trpD*、*trpC*、*trpB* 和 *trpA*，分别编码邻氨基苯甲酸合成酶、邻氨基苯甲酸焦磷酸转移酶、邻氨基苯甲酸异构酶、色氨酸合成酶和吲哚甘油-3-磷酸合成酶。*trpE* 基因是第一个被翻译的基因，和 *trpE* 紧邻的是 promoter、operator 和两个区［一个叫做前导区（leader）另一个叫做弱化子（attenuator），分别定名为 *trpL* 和 *trpa*（不是 *trpA*）］。色氨酸操纵子中产生阻遏物的基因是 *trpR*，*trpR* 距 *trp* 基因簇很远，*trp* 基因簇在 *E.coli* 染色体图上 25min 处而 *trpR* 则位于 90min 处。在位于 65min 处还有一个 *trpS*（色氨酸 tRNA 合成酶），*trpS* 和携带有色氨酸的 $tRNA^{Trp}$这两个因素也参与色氨酸操纵子的调控作用（图 9-11）。色氨酸操纵子的转录调控的特点是，除了阻遏操纵系统外，还有弱化系统。

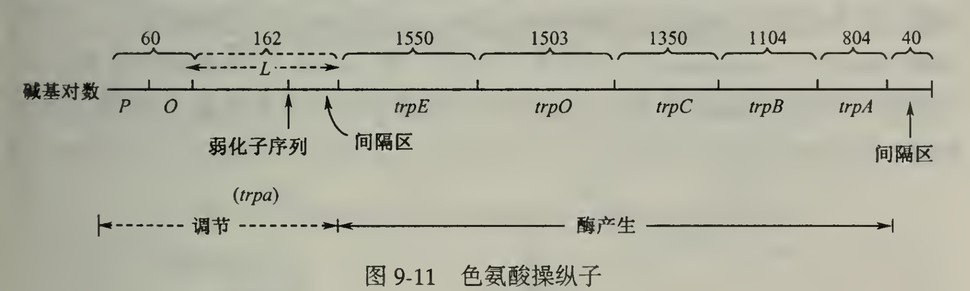


图 9-11　色氨酸操纵子

一、色氨酸操纵子的阻遏-操纵系统

trpR 基因突变型引起 trpmRNA 的组成型起始合成，这和在 *lac* 操纵子中一样，*trpR* 基因的蛋白产物常常称为 aporepressor，当没有色氨酸存在时，不能和 *trp* 的操纵区结合，当 *trp* 的 aporepressor 与色氨酸结合在一起时，成为有活性的阻遏物（repressor）才能与 *trp* 的操纵区结合。

aporepressor + operator ⟶ 有活性的启动子（active operator）　转录发生

aporepressor + tryptophan ⟶ 无活性的启动子（active repressor + inactive operator）　转录不发生

上两式表明，当色氨酸存在时，有活力的抑制物能抑制转录，当色氨酸的外部供应耗尽时，上面方程式中的平衡向左移动，操纵区不被抑制物占据，转录开始，这是最基本的开关（on-off）机制。

阻遏-操纵机制对色氨酸来说必须是一个充分的开关，主管转录是否启动，相当于色氨酸操纵子的一个粗调。而在色氨酸操纵子中对应于色氨酸生物合成还有另一个系统进行细调控，这种细调控主管已经启动的转录是否继续下去。在这个系统中，酶的浓度根据氨基酸的变化而变化。这个有效的调控是通过转录在达到第一个结构基因之前过早终止来实现的。由氨基酸的浓度来调节这种过早终止的频率。这些过程是如何完成的将在下面讨论。

二、弱化子和前导区

在 trpmRNA5′端的开头，在 *trpE* 基因的起始密码前有 162 个碱基。这段 mRNA 片段叫前导区。在前导区有一段碱基顺序(123～150)，如果这一段缺失了，引起 trp 的酶的合成提高 6 倍，无论是在一个阻遏的细胞里还是在一个组成型突变的细胞里都是这样（无论 $trpR^{+}$ 或 $trpR^{-}$）。这样说来，这段顺序(123～150）一定起调控作用。另外，当 mRNA 的合成起始以后，除非培养基中完全没有色氨酸，否则，转录总是在这个区域终止，产生一个仅有 140 核苷酸的 RNA 分子，使编码 trp 酶基因的转录短暂停止，这就是为什么这段缺失会导致 trp 酶的基因的表达提高。在这个区转录终止发生，并且这种发生是被调节的。这个区域就叫弱化子（attenuator)。引起终止的这一段 mRNA 的碱基顺序见图 9-12，此区有典型的终止区的特点，在 mRNA 中通过配对可以形成茎-环结构（stem-and-loop)。

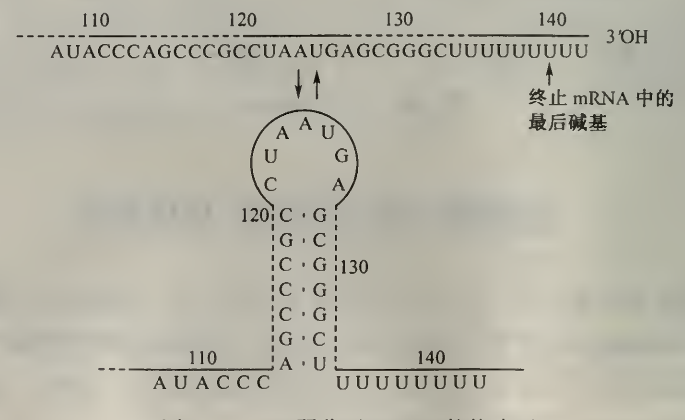


图 9-12　*trp* 弱化子 mRNA 的终止区

许多不同的实验指出，弱化需要负荷 trp 的 tRNA（$tRNA^{Trp}$）存在，由这一

点推测前导区的某一部分是被翻译的。对前导区的序列分析表明有起始密码AUG，在后边还有一个终止密码子UGA，如果翻译从AUG密码子开始，一个含14氨基酸的多肽应该被合成（图9-13）。不过在大肠杆菌细胞中至今还没有找到这个多肽，有可能这个多肽形成，但很容易分解，因此不易检测到。这个假想的肽称为前导肽（leader peptide）。有些证据表明这个14肽可能是被合成的，前导区内有SD顺序（1972年Shine-Dalgason发现的在mRNA起始密码子上游3～11碱基处的一个稳定顺序，与核糖体16S RNA的3′端互补，是核糖体结合部位，GAGG由于是Shine-Dalgason发现的所以叫SD顺序），这个SD顺序正好位于AUG的5′端。另外，前导肽编码区与*trpE*端融合，产生了一个N端与前导肽完全相同的融合蛋白。

假设的前导肽　　　　　　　　　　　　TrpE 蛋白质

Met Lys Ala Ile Phe Val Leu Lys Gly Trp Trp Arg Thr Ser Stop　　T G

pppAAG (23) AUG AAA GCA AUU UUC GUA CUG AAA GGU UGG UGG CGC ACU UCC UGA-(91) ··· AUG CAA ACA CAA

图9-13　*trp*前导区mRNA的碱基顺序

假设的前导肽中有两个并邻的色氨酸密码子

前导肽有一个很有意义的特点，就是在第10和第11位上有相邻的两个色氨酸密码子，这一点很重要，因为组氨酸操纵子中，也具有弱化子，也具有一个类似的能编码前导肽的碱基序列，此序列中含有7个相邻的his密码子。

三、mRNA的前导区全序列分析

mRNA的前导区已做了全序列分析，发现前导区的结构特点是4个部分可在不同条件下互相配对（图9-14），1区和2区，3区和4区形成发夹结构（或茎

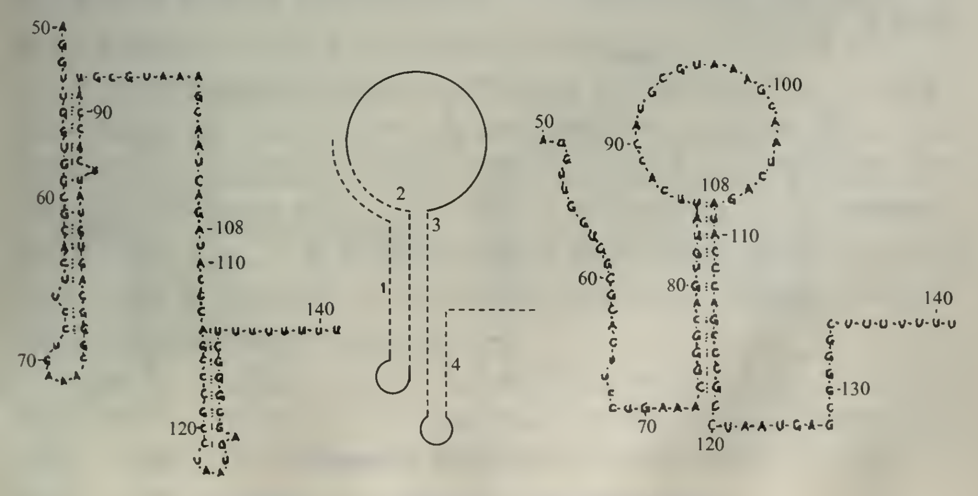


图9-14　trp前导区mRNA二级结构的几种形式

环结构），3 区和 4 区茎环后边紧接着 8 个 U。用 T1 RNase 处理（此酶不降解双链 RNA）表明，在提纯的色氨酸操纵子的转录产物中存在 1-2 和 3-4 茎环结构，而 3-4 茎环结构在终止识别区，这个区对于终止是必需的，在此区的几种突变都能在体内防止终止。在 1、2、3、4 区中，除了 1-2 和 3-4 配对外，在某种情况下 2-3 也可以配对，当 2-3 配对时，1-2，3-4 则不能配对形成茎环结构，这些因素共同构成了导致在 trp 操纵子的前导区中形成过早终止机制的理论。

四、弱化的机制

这个机制认为 mRNA 转录的终止是通过前导肽基因的翻译来调节的，因为在前导肽基因中有两个相邻的色氨酸密码子，所以翻译这个前导肽的能力必定对于 $tRNA^{Try}$（携带有色氨酸）的浓度是敏感的。在前导序列中，除 1 区与 2 区互补，2 区与 3 区互补外，3 区又和 4 区互补，这四个区在不同条件下配对情况不同，当培养基中色氨酸的浓度很低时，负载有色氨酸的 $tRNA^{Trp}$也就少，这样翻译通过两个并邻的色氨酸密码子的速度就会很慢。由于在原核生物中，转录和翻译是偶联的，即 mRNA 一边转录，核糖体一边结合上去翻译出多肽，在这种情况下，在 4 区被转录完成时，核糖体才进行到 1 区（或停留在两个相邻的 trp 密码子处），这时的前导区结构是 2-3 配对，3-4 不形成终止结构，所以转录可继续进行直到将色氨酸操纵子中的结构基因全部转录完。而当培养基中色氨酸浓度高时，核糖体可顺利通过两个相邻的色氨酸密码子，在 4 区被转录之前，核糖体就到达 2 区，这样使 2-3 不能配对，那 3-4 则可以配对形成茎环结构（终止子结构）使转录暂时停止，色氨酸操纵子中的结构基因不被转录，因而不再合成色氨酸（图 9-15）。实际上 RNA 聚合酶在弱化子处的活性依赖于前导肽翻译中核糖体所处的位置（图 9-16）。

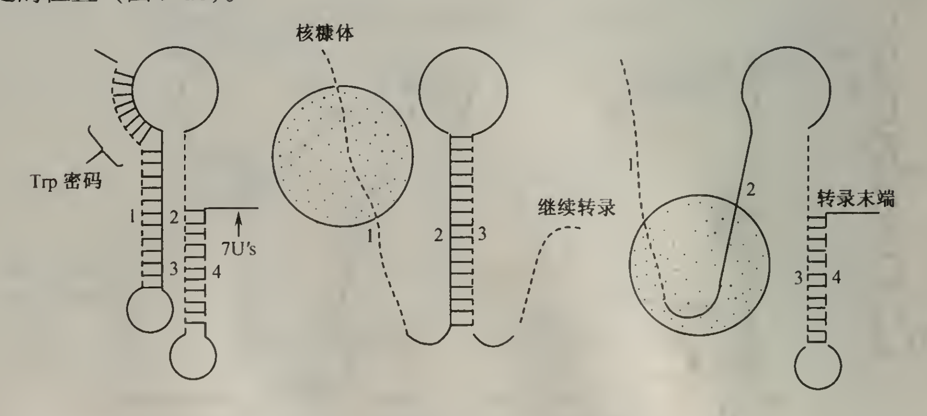


图 9-15　trp 操纵子中弱化机制模型

大肠杆菌和鼠伤寒沙门氏菌中已陆续发现不少操纵子都具有弱化现象。从前导区内密码顺序可以看出它们都具有假想的前导肽，并且在前导肽中都富含该操纵子合成的那种氨基酸，例如苯丙氨酸操纵子的前导肽中有六个苯丙氨酸（Phe），组氨酸操纵子的前导肽中有七个组氨酸（His）等。

色氨酸操纵子的一个重要特点是它具有弱化现象，那么上面讨论的有关色氨酸操纵子的基因表达的调节的理论是如何得的呢？下面我们将举几个具体的实验来说明。

五、色氨酸操纵子弱化机制的实验依据

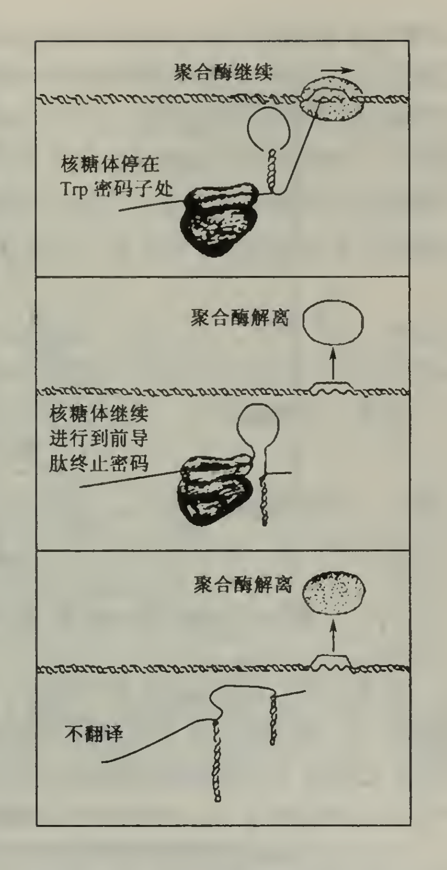


图 9-16　*E. coli* trp 操纵子前导区的转录与前导肽的翻译以及翻译中不同情况对结构基因转录的调控

1. 弱化作用的发现

早已知道，在大肠杆菌乳糖操纵子中的几个结构基因的表达是通过阻遏物作用在转录水平上进行调控的，在不同环境条件下，基因产物的浓度可以相差数千倍。在细菌的合成代谢中也发现过这种方式。例如色氨酸操纵子受阻遏蛋白的负控制作用的基因表达可下降为原来的 1/70。以后又发现在色氨酸操纵子中还有另一种在转录水平上的调控方式，它仅使基因表达下降为原来的 10%～12.5%，这一种在转录水平上的调控称为弱化。弱化作用最早是在色氨酸操纵子中发现的，发现了有些现象与以阻遏蛋白的阻遏作用为唯一调节机制的观点是不一致的。例如，一株色氨酰-tRNA 合成酶缺陷的大肠杆菌突变株 trpS5，不能完全阻遏色氨酸操纵子上编码邻氨基苯甲酸合成酶（ASase）的 *E* 基因的表达，即 trpS5 突变株在含有较高浓度色氨酸的培养基中，仍能使色氨酸操纵子的结构基因表达。

trpS5 是一个温度敏感突变株，它所编码的 trp-tRNA 合成酶只在 30℃ 时有活力，而在 42℃ 时便没有活力，也就是说在 30℃ 时才能使色氨酸转化为 trp-tRNA。比较野生型和突变型在 42℃ 和 30℃ 中，当供应给色氨酸时的 ASase 比活的大小，结果表明色氨酸-tRNA 具有调节功能（图 9-17）。

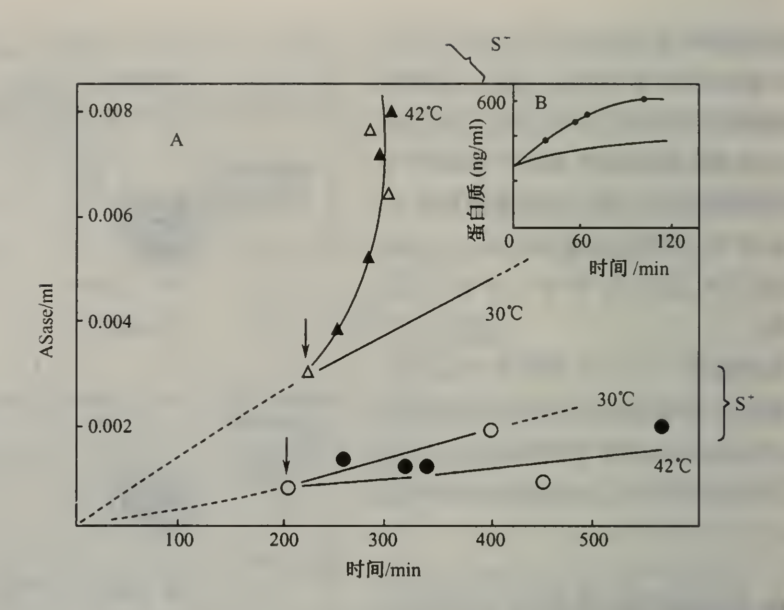


图 9-17　trpS⁻和 trpS⁺在 30℃和 42℃时其 ASase 的活力比较结果

从图上可以看出 trpS5 在 42℃时，trp 操纵子中基因表达增加（ASase 的比活高），在 30℃时，由于色氨酸乙酰化酶有活性，可以形成 trp-tRNA，因而 ASase 比活低（trp 操纵子结构基因表达低），而野生型的在两种温度下表达都差不多，这就说明了 trp-tRNA 对色氨酸操纵子结构基因的表达有调控作用。

另一突变株的获得及其特性的研究更充分说明了这种转录调控机制弱化现象的存在。在色氨酸操纵子突变株的筛选中，选到一株前导区直到 *D* 基因缺失的突变株（trpΔLD102），trpΔLD102 在补充了过量的色氨酸的培养基上，色氨酸合成酶（β-2）的活力比非缺失株大 8～10 倍。这种消阻遏作用不论是在 trpR⁺或 trpR⁻菌株中都存在。

表 9-2　大肠杆菌两菌株色氨酸合成酶活力比较

品　系	trpR⁺	trpR⁻
trpΔED_{53}（trpa⁺）	0.17	11.8
trpΔLD_{102}（trpa⁻）	1.82	100

从表 9-2 中数据可以清楚地看出前导区的调节作用。trpΔED53 中的 L 不缺失（弱化子存在），trpΔLD102 中 L 缺失（弱化子缺失），缺失前导区后的表达比有前导区的表达要高 8～10 倍多，说明色氨酸操纵子中的酶合成的调节除阻遏作用外，还受到前导区某种调节因素的影响，失去了这一因素就失去了调节作用，有关的酶便大量合成。

2. 弱化模型的遗传学证据

曾经发现过前导区发生突变的突变型，从表型变化上可以把它们分为两类：

一类突变使色氨酸操纵子表达下降，也就是有利于终止；另一类则使表达增加。也就是解除终止。trpL29 是增强终止的突变型，它的前导区内第 29 位核苷酸由原来的 G 变为 A，……AUG……→……AUA……，AUG 是前导肽的起始密码，变为 AUA 后便不能起始翻译，于是总是形成 1-2、3-4 稳定的二级结构，从而终止增强。这一突变型的邻氨基苯甲酸合成酶（ASase）活力低。并且对色氨酸不敏感，即对色氨酸饥饿不能做出解除终止的反应，这是由于前导肽根本不能翻译下去的缘故。

另一个增强终止的突变型是 trpL75，它的第 75 位上的 G 变成了 A。野生型 75 位上的 G 在 RNA 链的 2 区内，它与 RNA 链的 3 区内 118 位上的 C 可以形成氢键，当 G 变为 A 后便可减弱 2 区和 3 区成茎环能力，于是导致终止增强，trpL75 对色氨酸饥饿同样不能作出反应。

trpL29 和 trpL75 经再次突变可以得到回复子，回复子有两种类型：一类是真正的回复突变。例如 trpL29 中的 A 重新变为野生型中的 G；另一类则是在另外位置上的核苷酸又发生了一次改变，这类突变都是发生在 3 区和 4 区，或者 3′端的一串 U 上，例如双突变型 trpLG→C trpL117ΔC 和 trpL29G→A trpL131G→C 等，前者是 117 位上缺失了 C，后者是 131 位上 G→C，这两种变化都不利于 3 区和 4 区形成稳定的二级结构，从而使终止解除。这两株回复型的 ASase 活力都比 trpL29 为高，此外还找到一株 ΔtrpLC1419，它缺失了 3′端的四个 U，也导致

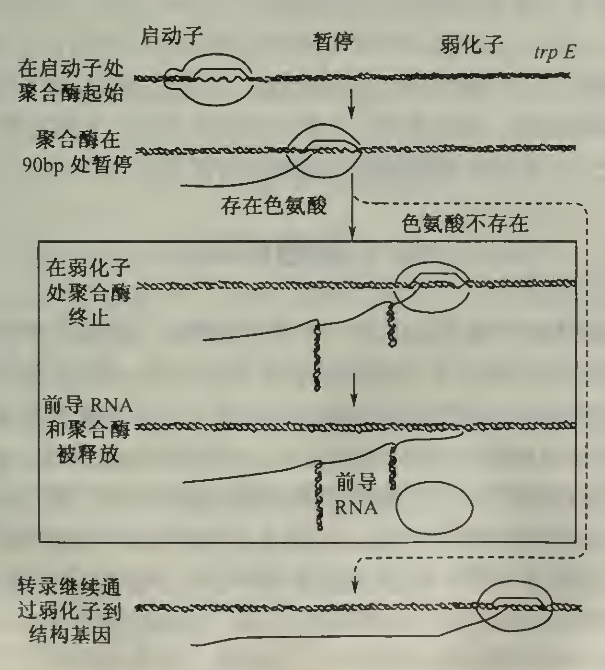


图 9-18　色氨酸操纵子转录调控（包括阻遏-操纵和弱化子）

终止解除，说明3′端U对转录终止起一定作用。细菌通过弱化作用辅助阻遏作用的不足，阻遏作用只能使转录不起始，对于已经起始的转录；则只能通过弱化作用而使之中途停顿下来。阻遏作用的讯号是细胞内色氨酸的多少；弱化作用的讯号则是细胞内载有色氨酸的tRNA的多少。它通过前导肽的翻译来控制转录的进行，在细菌细胞内这两种作用相辅相成，体现着生物体内精细的调控作用（图9-18）。

第四节 λ噬菌体基因表达的调节

前面讨论的启动子和终止子转录的控制可以说是一种固定的控制装置，即转录的起始和终止要求DNA上有某种特定的结构，而这种结构的功能不受环境条件的影响。弱化子则既有像终止子那样固定的结构，又由于生理条件的变化而改变结构，从而行使它的转录调节功能。乳糖操纵子受阻遏物的调节是环境因素在转录水平上调节的典型例子。λ噬菌体的转录则是更为复杂的调控体系，因为它和λ噬菌体的裂解途径和溶源化途径的选择密切相关，而涉及更多基因的转录，它也是到现在为止从基因转录的调控角度研究得最为深入的体系。它与其他调控体系最大的区别是：①抗终止，在λ噬菌体中发现*N*基因的产物N蛋白能解除终止子的终止作用，这一作用称为抗终止作用，λ噬菌体基因向左和向右转录的两个操纵子上各有一些能与PN作用的顺序，即PN作用部位（N-utilization site, Nut），在右操纵子上称为NutR，在左操纵子上的称NutL，这两个Nut突变后，PN不能抗终止。另外，*Q*基因产物Q蛋白也起抗终止作用；②λ噬菌体处于溶源循环时，主要是有*CI*基因产生λ阻遏物，这种阻遏物结合于左、右操纵区，阻止λ其他基因的转录，同时根据CI蛋白的浓度不同，它除了阻遏λ其他基因转录外，还阻止*CI*基因本身的转录，这称为自调节。

一、λ噬菌体简介

λ噬菌体是以大肠杆菌为宿主的一种温和噬菌体，感染大肠杆菌以后，将其基因组整合到细菌染色体上成为细菌染色体的一部分，随着染色体的复制而复制。这种整合有噬菌体基因组的细菌称为溶源菌，这一过程称为溶源化途径，溶源化细菌一般不被同种噬菌体再行感染，这一现象称为免疫性。λ噬菌体感染宿主后也可以不进行溶源化途径而利用细菌细胞内的养料进行繁殖，最终使宿主裂解而死亡，这一过程称为裂解循环。溶源化的细菌受某些外界物理或化学因素（如紫外线、丝裂霉素C等）的作用原噬菌体便脱离细菌染色体而进行自主复制，于是细菌裂解，游离出大量的噬菌体，这一过程称为诱导。

λ噬菌体是一种中等大小的噬菌体，它的基因组是由双链线状DNA分子组成。分子质量为32.3×10^6 Da，由48 502bp组成，排列顺序已全部定出。两条

单链的 5′端有突出的 12 个碱基，具有回文对称特点，12 个碱基中有 10 个是 G 或 C，可以牢固地配对形成环状（图 9-19），因此这 12 个碱基称为黏性末端，这配对的 12bp 部分称为黏性位点。从 λ 噬菌体提取 DNA，用限制性酶 *Eco*RI 消化可得六个片段，如若先经退火后再行酶切则只得五个片段。可见 λ DNA 在噬菌体内是以线形存在的，感染细菌后便行环化，然后再进行各种生理活动。λ 噬菌体的黏性末端对于噬菌体的感染活性很重要，当把它切平后，就失去感染性。

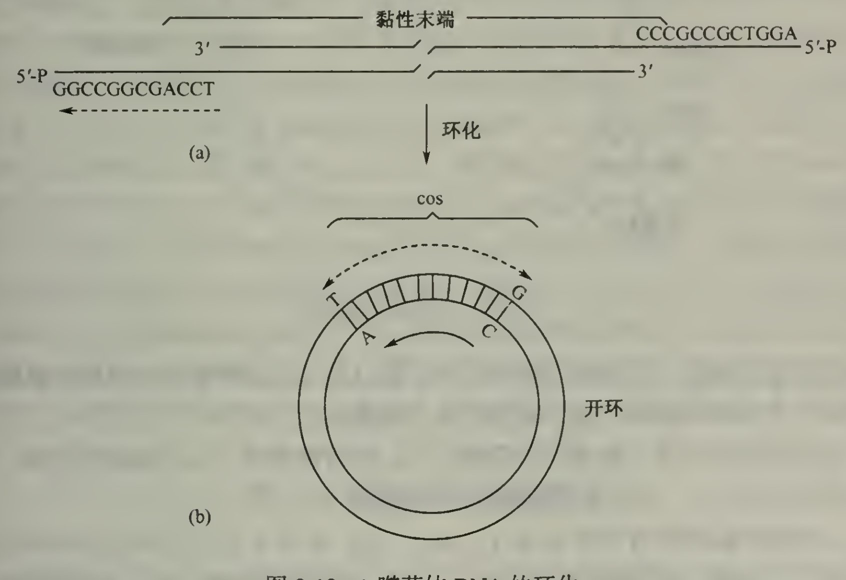


图 9-19　λ 噬菌体 DNA 的环化

二、λ 噬菌体基因组

λ 噬菌体基因组上大约有 35～40 个为蛋白质编码的基因，另外还有一些识别位点。大多数蛋白已被纯化，有些已被测出顺序，λ 基因调控通过转录进行，与调控有关的基因在基因组上比较集中。在大肠杆菌中一个操纵子便成为一个转录单位，每一个操纵子中包括功能上密切相关的几个基因；λ 的一个转录单位则包括功能上并不直接相关的更多基因。λ 的转录调控不但包括起始和终止，而且还包括和裂解途径或溶源途径有关的两种途径的选择（图 9-20）。

从图 9-20 中可以看到 λ 基因组分四大基因簇，与调节控制有关的调节区（regulation）；与重组有关的区域（recombination）；与复制有关的复制区（replication）和结构基因区（包括：头、尾、装配和裂解有关的基因）。

λ 基因组上有三个调节基因，*N*、*Cro* 和 *Q*，N 蛋白抗终止子 tL1、tR1 和 tR2；Q 蛋白抗终止子 tR4；Cro 蛋白对 *CI* 和 *Cro* 基因的转录起负调控作用。

与溶源化有关的有三个基因，*CI*、*CII* 和 *CIII*，CIII 蛋白使 CII 蛋白稳定，

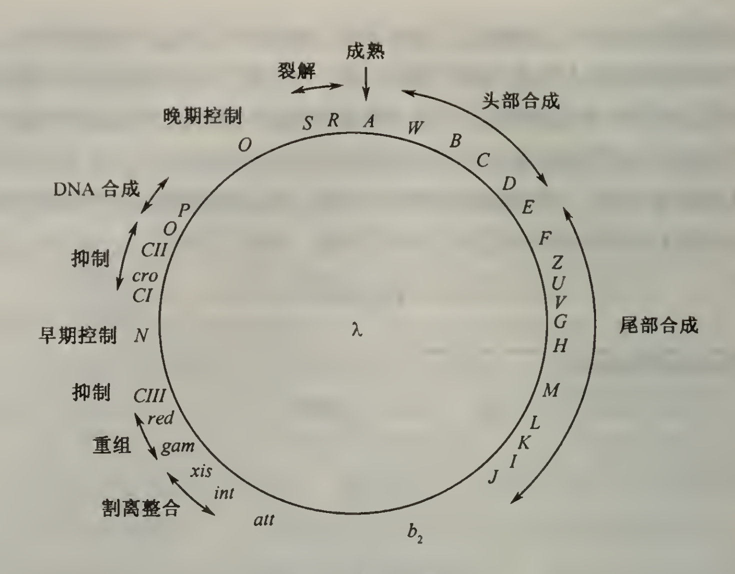


图 9-20　λ 噬菌体基因组

CII/CIII 蛋白能使 *CI* 基因启始转录，CI 蛋白是 λ 阻遏物能使 λ 的复制成熟和裂解等有关基因都不转录，使 λ 噬菌体处于溶源状态。

重组区基因主管 λ 的整合和割离，*xis* 编码割离酶，*int* 编码整合酶，整合酶作用位点在 *att*，作用是打断此处的 DNA 链。

λ 噬菌体基因组有六个启动子，P_L、P_R1 分别负责启动向左向右的转录，P_E、P_M 是转录 *CI* 基因的两个启动子，P_E 主管建立溶源，P_M 主管维持溶源，P_E 的启动需要 CII、CIII 蛋白，P_I 启动转录 *int* 基因，P_I 的起始也需要 CII/CIII 蛋白，P_R2（或 P_R'）是 λ 最强的启动子，负责后期一切基因的转录。P_L、P_E、P_M、P_I 是启动左向转录，而 P_R1、P_R2 则启动右向转录（图 9-21）。

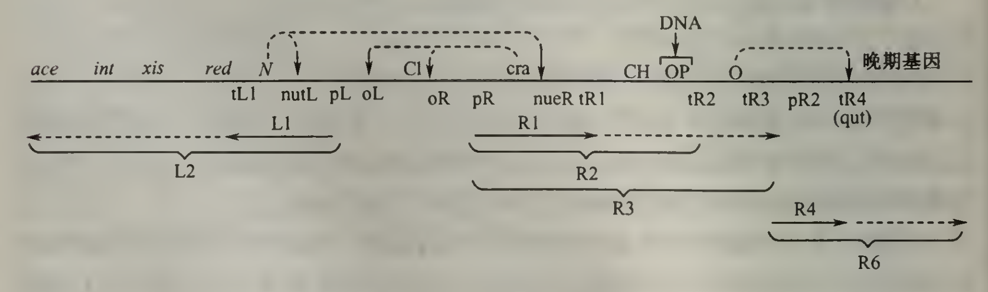


图 9-21　λ 噬菌体调控区

三、λ 噬菌体感染宿主后的转录次序

λ 噬菌体没有编码 RNA 聚合酶的基因，它借用宿主的酶合成 RNA。在 λ 噬

菌体的转录中，宿主的 RNA 聚合酶也被修饰了，这种修饰不是为了去识别噬菌体的启动子，而是修饰后使 RNA 聚合酶能抗终止，事实上 λ 噬菌体晚期 mRNA 的合成被延迟的主要原因就是转录翻译一种抗终止蛋白需要时间（λ 噬菌体裂解循环需要 45min，而 T7 只需要 22～25min，λ 这样长的生活周期反映了 λ 噬菌体更换生活周期所需要的复杂的调节途径），λ 基因转录从两条链向不同的方向进行，向左转录的一条链为 L 链，向右转录的一条链称为 R 链。λ 基因组由若干个操纵子所组成，根据各个操纵子转录的先后可把转录分为三个期。L1 和 R1 为早期转录产物，它们分别由左右两个操纵子上的左启动子（P_L）和右启动子（P_R）所启动，*R1* 和 *L1* 的 mRNA 主要分别译出调节蛋白 Cro 和由 *N* 基因编码的 N 蛋白（PN）。*R1* 和 *L1* 分别在终止子 tR1 和 tL1 上停止转录，但由于有 PN 的出现，它的抗终止功能可以使 *R1* 和 *L1* 继续转录，形成 *R2* 和 *L2* 的 mRNA，翻译后分别产生功能上与 CI 蛋白密切相关的 CII 和 CIII 蛋白以及 DNA 复制必需的 O 及 P 蛋白，转录继续进行达到终止子 tR2 时也并不终止。因为 PN 蛋白对 tR2 也有抗终止作用，转录最后进行到 tR3，翻译出 Q 蛋白，R2、L2 和 R3 便是这一阶段——晚早期转录的转录产物。第三阶段即晚期转录，从启动子 P_R2 开始转录首先到达终止子 tR4，由于 Q 蛋白对它的抗终止作用，转录继续进行。使基因组上头部蛋白，尾部蛋白，与装配、裂解有关的基因全部表达。

四、λ 噬菌体的调控区

λ 噬菌体基因组内与调控有关的基因或位点集中在 CIII 和 CII 之间范围内，这一区域也称为调控区。调控区内共有四个启动子：右启动子 P_R、左启动子 P_L 和另外两个启动向左转录的启动子 P_E 和 P_M。

从 P_R 转录并经翻译得到 Cro 蛋白和 CII 蛋白是决定溶源化和裂解的关键。从 P_L 转录并经翻译得到 P_N 和 CIII 蛋白，P_N 起抗终止作用，CII 和 CIII 蛋白起着激活启动子 P_E 的起始转录作用。λ 感染细菌之后所形成的 CII 和 CIII 激活从 P_E 开始的向左转录（包括 *CI* 基因的转录），*CI* 基因的产物是阻遏物蛋白，它阻止 λ 的复制。λ 整合到宿主细胞的染色体上之后，仍然合成少量的阻遏物，这一转录由启动子 P_M 所启动，事实上这是原噬菌体的唯一转录产物。P_E 即建立溶源化的意思。P_M 即维持溶源化的意思。由 P_E 启动的转录得到的 CI 蛋白量要比由 P_M 启动得到的大 5～10 倍。这可能是由于 P_M 中的 Pribnow 顺序与起始密码相距很近（这种情况在一般的 mRNA 中很少见），因为缺少了 SD 顺序，从而翻译的效率下降。由 P_E 起始转录的 mRNA 较长，含有 SD 顺序，所以能翻译出更多的 CI 产物。这种调节是在翻译水平上进行的。

五、溶源化的遗传控制及λ阻遏物的发现

我们已经知道λ噬菌体是温和性噬菌体，能使宿主形成溶源菌，溶源菌有免疫性，即溶源菌有防止同种噬菌体超感染的能力，当溶源菌失去了原噬菌体后(即治愈了的细胞）对以前它所携带的噬菌体是敏感的，说明溶源菌的免疫性必定是由原噬菌体引起的。为了证明这一点，Jacob 和 Wollman 用 Hfr（λ）和 F^- 进行杂交实验，看有无λ噬菌体释放。

①Hfr（λ）$\times F^-$ 释放 100～200 个λ噬菌体颗粒，这一现象称为结合子诱导(zygotic induction)。②Hfr（λ）$\times F^-$（λ）不释放噬菌体；Hfr $\times F^-$（λ）不释放噬菌体。从以上两种情况中可以看出，只要受体菌是溶源菌，不管供体菌是否是溶源菌都不产生结合子诱导，而只有当受体菌是非溶源菌而供体菌是溶源菌时才产生结合子诱导。这说明受体菌中的原噬菌体不仅对游离的λ的超感染有免疫性，而且对 Hfr 供体菌转移过来的原噬菌体也有免疫性。因此推测溶源菌的免疫性是由于原噬菌体产生了一种免疫抑制物。此抑制物可抑制λ噬菌体的复制和λ其他基因的转录，因此认为温和性噬菌体的本质在于它本身合成自我抑制物，而原噬菌体的诱导是细胞内的抑制物降低到一定水平的结果。因此那些能妨碍抑制物合成或破坏抑制物功能的因素即能够起诱导作用。以上实验证明免疫抑制物是细胞质因子，那么这个免疫抑制物是由什么基因控制呢？要证明这个问题，需要选到不产生抑制物的突变株。

早期对λ噬菌体的研究主要是关于噬菌斑类型的突变（plaque type mutation)。野生型λ噬菌体感染宿主细菌后形成的噬菌斑是混浊的，发现有一种λ噬菌体的突变株形成的噬菌斑是清晰的，以后又发现几种清晰程度不同的λ噬菌体突变型，便分别用 CI、CII、CIII 来表示，经测定，它们的溶源化频率和野生型不相同：

菌种	溶源化频率
+	约 4×10^{-1}
CI	约 1×10^{-5}
CII	10^{-4}～10^{-5}
CIII	10^{-2}～10^{-1}

CI 基因突变的分子基础是什么？以下几方面的事实说明 *CI* 基因编码一种阻遏蛋白，它的作用正像乳糖操纵子的 *lac I* 基因产物的作用。

1) *CI* 基因突变后使宿主细菌不被溶源化而发生裂解，说明 *CI* 的作用是阻遏λ噬菌体的活动。*CI* 基因和大肠杆菌的 *lac* Ⅰ相似。

2) CI 的作用不但阻止已溶源化的λ原噬菌体的活动，也使超感染的λ噬菌体不能复制（这就是免疫性的基础)。这也和 *lac I* 产物相似，即顺位反位都是显

性，因此也可以设想 *CI* 基因也编码一种蛋白质。温度敏感突变型的存在也同样说明 *CI* 基因产物是一种蛋白质。

3) 也像 *lac I* 基因产物一样，*CI* 基因产物作用在 λ 噬菌体的特定位置上。下列事实可以说明这一点：溶源化细菌对 λ 噬菌体具有免疫性，但是 λ imm434、λ imm21 等杂种噬菌体感染时，这些噬菌体能在 λ 的溶源菌中复制繁殖。这是因为 λ imm434 和 λ imm21 的免疫区已被 434 和 21 的免疫区所代替，所以 *CI* 的产物不能对它们发生阻遏效应而得以复制繁殖，可见免疫区便是 CI 蛋白作用的特定位置。

六、λ 噬菌体的操纵区和启动子结构

λ 噬菌体的基因组上的左右操纵子也像大肠杆菌的 *lac* 操纵子一样，在结构基因之前有启动子 *P* 和操纵区 *O*，并且 *P* 与 *O* 之间也有重叠。λ 操纵区结构与 *lac* 操纵区有不同之处，λ 操纵区有三个能和阻遏物结合的位点。即右操纵区中的 O_R3、O_R2、O_R1 以及它们相对应的左操纵区中的 O_L3、O_L2、O_L1，如图 9-22所示。

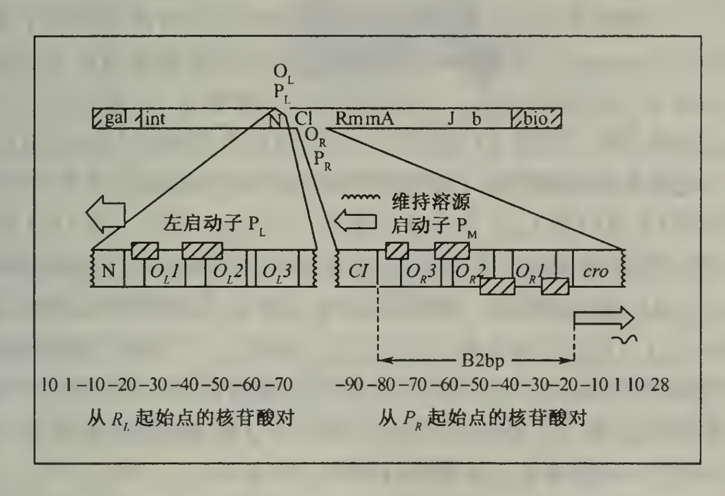


图 9-22　λ 噬菌体左右操纵区的结构

每个结合位点都由 17bp 组成，每个位点之间相隔 6～7bp，大多数是 A/T，两个操纵区共六个结合位点，这六个结合位点都表现有一定程度的回文对称性。曾经筛选到一些 λ 的烈性突变型，能在 λ 的溶源菌中生长繁殖，这些点突变大多减弱了结合位点的回文对称性。看来操纵区的回文对称性与阻遏物的结合是有关的。乳糖操纵子操纵区也有一定程度的回文对称性；乳糖操纵子或半乳糖操纵子上 CAP 结合位点等也有一定程度的对称性。下列是一个 17bp 低聚物，是参照六

个结合位点人工合成的，它的特点是和任何一个位点都不一样，但又呈绝对的回文对称性，它和λ阻遏蛋白亲和力最大。

TATCACCGCCGGTGATA

ATAGTGGCGGCCACTAT

可见对称程度越大则亲和力越大。但是天然的结构不是绝对的对称形式，可能是因为亲和力过大，不利于解离，不能起到良好的调节功能。

从λ噬菌体的DNA顺序上看，它的启动子和操纵区有重叠（图9-22），P_R 的Pribnow顺序和－35顺序与 O_R1、O_R2 重叠，P_L 的也一样，*CI* 基因的启动子 P_M 的Pribnow顺序和－35顺序则和 O_R3、O_R2 有重叠，这些特点就构成了λ噬菌体对其发育过程调节的分子基础。

七、CI 蛋白和 Cro 蛋白

CI蛋白也称阻遏蛋白，是一种酸性蛋白，由236个氨基酸组成，CI阻遏蛋白由两个相同的亚基组成。它的碱性氨基酸只占10%，大部分集中在N端，N端26个氨基酸中就有9个是赖氨酸或精氨酸，占N端氨基酸的35%。这样的结构究竟有什么生物学意义呢？遗传学分析和生物化学分析都说明每个亚基由两个功能不同的区（domain）组成，N端具有与DNA上操纵基因结合的功能，C端具有自身聚合为二聚体的功能。C端的聚合功能解释了CI蛋白与三个结合位点之间结合的协同作用，即当CI蛋白与一个位点结合后使它与相邻位点的结合更为容易。用足迹法可以测定各个操纵基因结合位点与CI蛋白结合时的相对浓度。实验结果发现CI蛋白与 O_R1 结合能力最大，与 O_R2 次之，与 O_R3 结合能力最弱。当CI蛋白浓度较低时，它首先与 O_R1 结合（同样也与 O_L1 结合），占据了 P_R（和 P_L）的Pribnow顺序，因而阻止 P_R（和 P_L）启动转录，当CI蛋白浓度较高时除与 O_R1（O_L1）结合外，还与 O_R2（和 O_L2）结合，此时能促进 P_M 启动的转录（转录 *CI* 基因），而当CI蛋白浓度很高时，CI蛋白除与 O_R1、O_R2 结合外，还与 O_R3 结合，这时CI蛋白占据了 P_M 的Pribnow顺序和－35区，因此阻止 *CI* 基因本身的转录，这就是自调控（图9-23）。

Cro蛋白于1976年被分离获得，两年后测定了它的一级结构，它由66个氨基酸组成，Cro蛋白与CI蛋白在一级结构上无明显的同源性，也不分为两个功能区，可是功能上与CI蛋白非常相似，也是以二聚体形式与操纵区结合，Cro蛋白和三个结合位点的结合能力也不同，它和三个结合位点的结合能力与CI蛋白与三个结合位点的结合能力恰好相反，Cro蛋白与 O_R3 结合能力最强，与 O_R2 次之，与 O_R1 结合能力最小。Cro蛋白浓度低时它首先同 O_R3 结合阻止 *CI* 基因转录，浓度较高时与 O_R2 结合除阻止 *CI* 基因转录外，对 P_R 也有阻遏作用，浓度高时，与 O_R1 也能结合，阻止 P_M 和 P_R 启动转录。

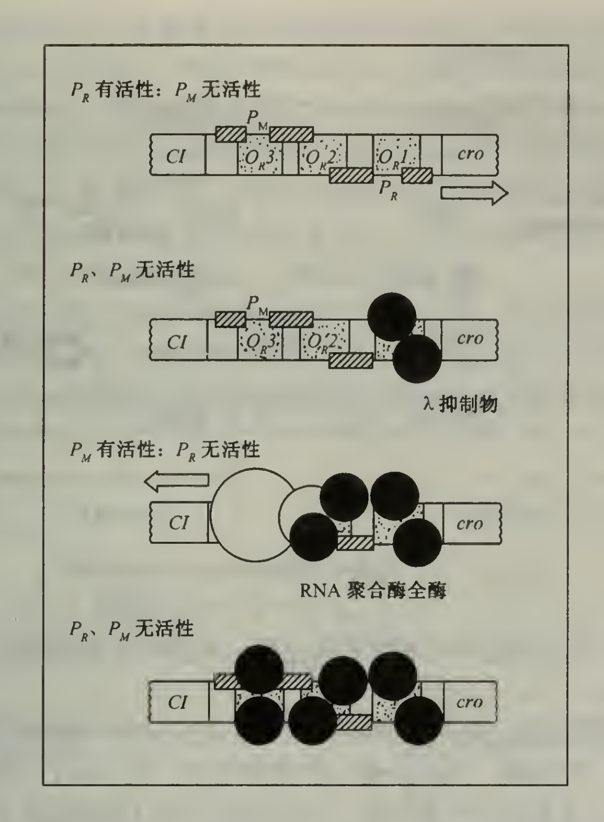


图 9-23　λ 抑制物与右操纵区结合对 P_R、P_M 的活性的正负调控

应用融合操纵子的技术证明了 CI 蛋白对 P_R（产生 Cro 蛋白）起负控制作用；对 *CI* 本身有正控制作用也有负控制作用。Cro 蛋白对 *CI*（P_M）和 P_R（*Cro*）都起负控制作用。

λ 噬菌体行溶源循环时，λDNA 整合到宿主染色体上，随染色体复制而复制，当溶源菌受紫外线或丝裂霉素 C 处理后，激活 recA 蛋白，recA 蛋白具有蛋白水解酶活性，分解 CI 蛋白。将 CI 蛋白在第 111 和 112 氨基酸之间打断，C 末端先从 DNA 上掉下来，从而 N 端也掉下来，于是 RNA 聚合酶便有机会与 *Cro* 基因的启动子结合，使转录向裂解方向进行。

λ 噬菌体感染大肠杆菌后，转录首先从 P_L 和 P_R 启动，产生 N 蛋白和 Cro 蛋白，N 蛋白可抗从 P_L 和 P_R 起始的终止［图 9-24（a)］。N 蛋白的浓度足够时，可以进行合成更长的 RNA［图 9-24（b)］。

突变研究表明，N 蛋白抗终止的作用专一序列（被称为 nut N-utilization)，位于启动子和终止子之间。

分离到一些 *E. coli* 突变株，不能提供由 N 蛋白引起的抗终止，进而限制 λ

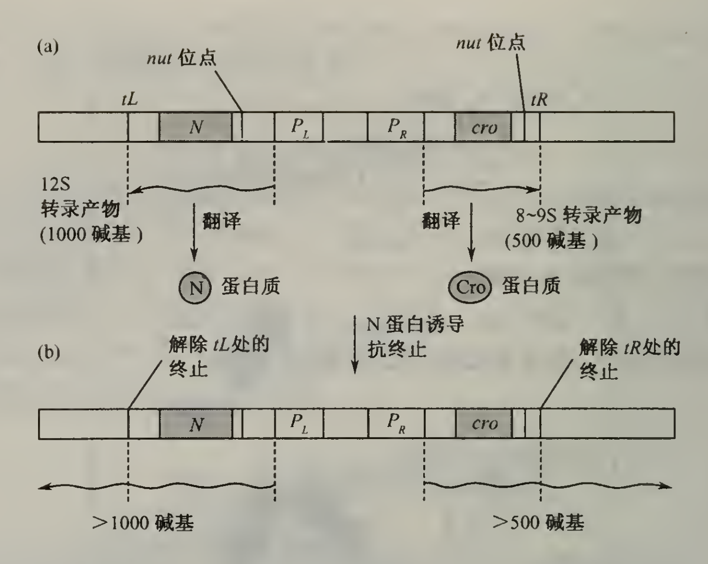


图 9-24　从 λ 噬菌体基因组的 P_L 和 P_R 的转录

噬菌体的感染，从中鉴定出几种细胞蛋白质，或者叫 N 利用底物（Nus），是 N 功能所必需的。

这些蛋白质定名为 Nus A、Nus B、Nus E 和 Nus G，这些蛋白在被感染的细胞中具有其他一些功能。Nus A 是一种延伸因子，而 Nus E（也称为 S10）是 *E. coli* 核糖体小亚基的一个蛋白质，这些细胞内蛋白质的功能共同作用防止核糖体 RNA 基因转录的终止。

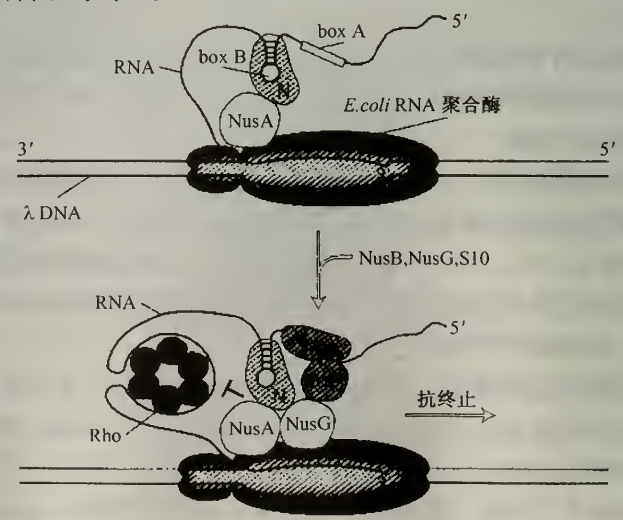


图 9-25　由 λ-噬菌体 N 蛋白和 *E. coli* 细胞蛋白的抗终止

蛋白质结合研究表明，N 蛋白的功能是结合在新生 RNA Nut 位点上，该 Nut 位点含两个 N 结合域：一个茎-环叫 box B，它与 N 相互作用，另一个 12 碱基的线性顺序叫 box A，它与细胞内蛋白 Nus B 和 S10 相互作用。N 介导的终止模型见图 9-25。N 蛋白结合到新生 RNA 上 nut 的 box B，然后与和 RNA 聚合酶复合在一起的 Nus A 相互作用。一旦 N 和 Nus A 相互作用，Nus B、Nus G 和 S10 迅速结合，产生一个稳定的抗终止复合物。该复合物可沿 DNA 移动数千 bp，抗依赖于 ρ 和不依赖于 ρ 的终止，使转录继续进行。

精致的抗终止机制，雇用一个多蛋白复合物，允许 λ 噬菌体调节从 P_L 和 P_R 转录的基因的暂时表达。这种调控机制和其他一些机制一起允许 λ 经历裂解或溶源感染（图 9-26）。这个例子也阐明基因表达可通过控制转录单位的过早终止来调节。

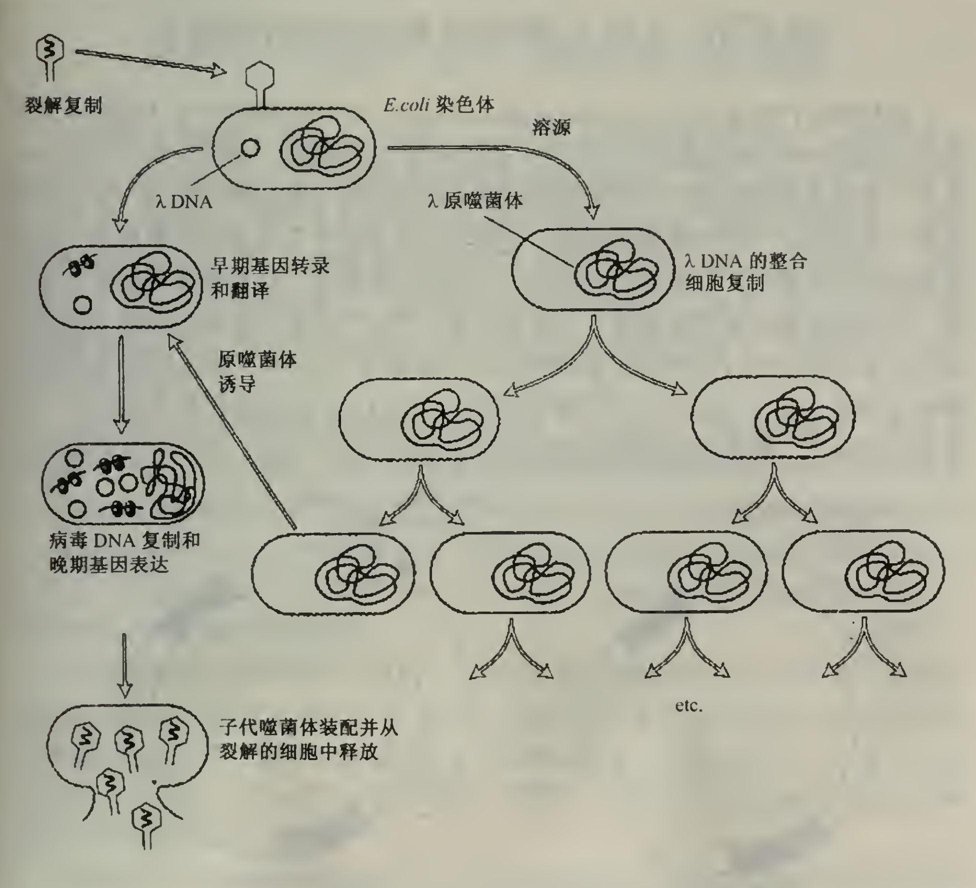


图 9-26　λ 噬菌感染 *E.coli* 后经历裂解复制或溶源化

Cro 蛋白可和左右操纵区结合（由于浓度低只结合于 O_R3）对 P_L 和 P_R 有弱

的阻遏作用，由 N 蛋白的作用使转录继续前进产生 CII、CIII、O、P、Q 蛋白，CII/CIII 能激活 P_E 和 P_I 启动子，转录后产生 CI 蛋白和整合酶，而 Q 蛋白量还很少，不足以启动 P_R2 完成晚期转录，O、P 蛋白的量也只能供 λDNA 复制一次。由于 P_E 转录产生 CI 蛋白结合于左右操纵区阻止 λ 其他基因转录，P_I 转录产生的整合酶使 λ 基因组整合到宿主染色体上而建立溶源。溶源建立之后，*CI* 基因从 P_M 开始转录，通过 CI 蛋白对它自身的自调控作用，从而保持细胞内的 CI 蛋白浓度在 10～20 个分子，使其维持溶源状态。

当环境因素及宿主的遗传背景或目前尚不清楚的因素使 CII 蛋白比 Cro 蛋白相对浓度低时，Cro 蛋白阻止 *CI* 转录，不产生 λ 阻遏物，使 λ 的其他基因可以转录，当产生的 Q 蛋白的量达到使 P_R2 转录完成时，进行裂解循环途径。

第五节　DNA 重排对基因表达的调节

鼠伤寒沙门氏菌（*Salmonella typhimurium*）有两种鞭毛抗原（即两种血清型），这是由两个基因编码的。这两个基因叫做 *H1* 和 *H2*，最早发现（1922 年）由单一菌落来的鼠伤寒沙门氏菌的群体常能被两种鞭毛抗血清所凝集。从中分离的单个菌落能被一种抗血清所凝集，可是在培养过程中又出现少数能被另一种抗血清所凝集的菌落。后者在培养过程中又出现少数前者。这种现象叫做鞭毛相转变，它不能用基因突变来解释，主要因为每一种沙门氏菌只在某两种鞭毛抗原类型中发生来回变化（而沙门氏菌的鞭毛抗原类型有十几种），而这种变化的发生频率又很高（10^{-5}～10^{-3}，自然突变率在 10^{-8}左右），遗传学研究告诉我们，鞭毛基因有两个，鞭毛抗原类型就是由这两个鞭毛抗原决定的。所以这些现象可以

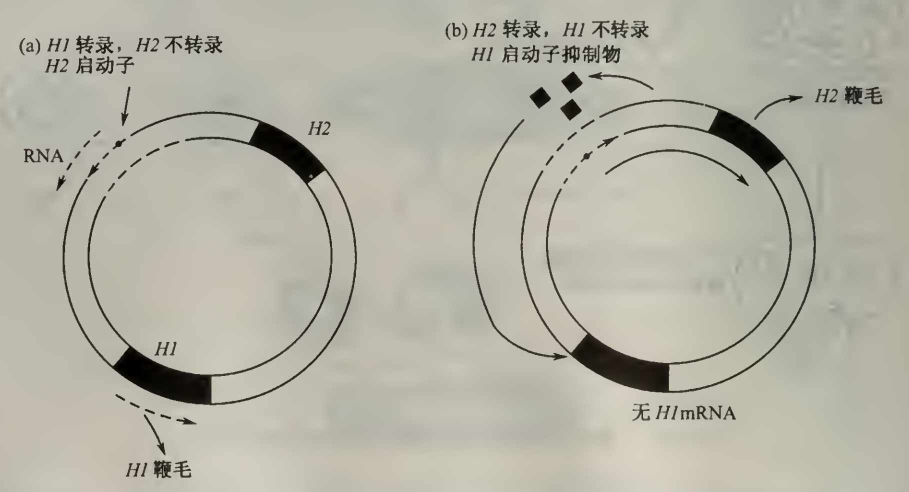


图 9-27　两种鞭毛抗原基因的调节

归纳为这样两个问题：①一个细胞中的两个基因任何时候只有一个表达；②是什么因素使一个基因表达，又使另一个基因不表达呢？

H1 基因和 *H2* 基因在染色体上相距很远，这就意味着必定有一种可扩散的产物在基因开关中起作用，*H1* 基因与其他正常基因一样，但 *H2* 基因有两种形式，*H2* 基因的一端有一段 970bp 顺序，这段 970bp 有三个特点：可以倒位；可以产生抑制 *H1* 转录的产物；带有 *H2* 的启动子（图 9-27）。

从图 9-27 中可以看到，当 *H2* 一端的 970bp 以一种方向存在时，*H2* 转录，*H2* 的 mRNA 不但编码 *H2* 的抗原而且编码 *H1* 的抑制物，因此当 *H2* 转录时，*H1* 就不被转录。如果 970bp 倒位，则 *H2* 不转录，*H1* 可以被转录。如何证明开动 *H1* 还是 *H2* 是由于倒位引起的？提取 H1 鞭毛菌 DNA 和 H2 鞭毛菌的 DNA—酶切—变性—复性—电镜观察可以证明。

第六节　西格马因子对基因表达的调控

通过替换西格马（sigma）因子（σ）可以控制转录的起始。大肠杆菌的 RNA 聚合酶只有一种，同一种酶可以转录所有的基因。这些基因是在不同场合以不同水平表达的，所以大肠杆菌 RNA 聚合酶必须具有广泛的识别范围。RNA 聚合酶在特定启动子上的转录活性似乎由一系列的辅助因子的循环来调控，而酶本身的亚基并不发生任何变化。这种假设并不被所有的人接受，因为允许 RNA 聚合酶识别某专一启动子的变化将可以防止它识别其他启动子，这将降低 RNA 聚合酶的灵活性。

RNA 聚合酶全酶中的核心酶主管链延伸，而 σ 因子主管位点选择。那么是否存在一种以上的 σ 因子，每一种特异性地识别不同类型启动子中的一种？在正常生长的大肠杆菌中只有一种 σ 因子。枯草杆菌中含有多种特性稍微不同的 σ 因子。在一定的环境中，当细胞的生活方式发生强烈变化时，大量基因表达的开关被打开，在这种情况下，枯草杆菌 RNA 聚合酶上原有的 σ 因子被具有识别不同特性启动子的 σ 因子全盘替换。

枯草杆菌在正常的营养生长时，其 RNA 聚合酶的结构与大肠杆菌的相同即 $\alpha_2\beta\beta'\sigma$。其 σ 因子的分子质量为 55 000U，称为 σ^{55}，像大肠杆菌中一样，以同样的标准识别启动子。酶的区别在于它含不同的 σ 因子，其数目很少，这些 σ 因子用 σ^{00}表示，“00”指示这种 σ 因子的分子质量，这些改变了的酶识别与主要酶 $\alpha_2\beta\beta'\sigma^{55}$不同的启动子。

在某种状态如噬菌体感染期间，σ 因子的替换可以引起从一套基因的表达转换成另一套基因的表达。裂解感染是噬菌体接管一个细菌并摧毁它，最后产生更多的噬菌体的过程。在这种情况下，噬菌体的发育包括转录途径的转变。在大肠杆菌中，通过合成由噬菌体编码的 RNA 聚合酶或者合成由噬菌体编码的辅助因

子控制大肠杆菌的 RNA 聚合酶来完成这种转变。在枯草杆菌被噬菌体 SPO1 感染期间，产生两种新的 σ 因子。SPO1 的感染史经过 3 个阶段的基因表达。刚感染时，噬菌体的早期基因被转录；感染 4～5min 后，早期基因停止转录而中期基因开始转录；当感染 8～12min 时，中期基因的转录被晚期基因的转录所代替。

早期基因的转录由宿主的全酶来进行，这些基因的启动区与宿主的基本无区别，它们的启动子都具有被 RNA 聚合酶 $\alpha_2\beta\beta'\sigma^{55}$识别的固有能力。

早期基因的转录转变到中期、晚期基因的转录需要噬菌体基因的表达，三个调节基因分别定为 *28*、*33* 和 *34*，它们控制转录的进程，如图 9-28 所示。调节的途径造成一个级联，在这个过程中，宿主的酶转录早期基因，早期基因产物对中期转录是必需的，而两个中期基因编码的产物对于晚期基因转录又是需要的。

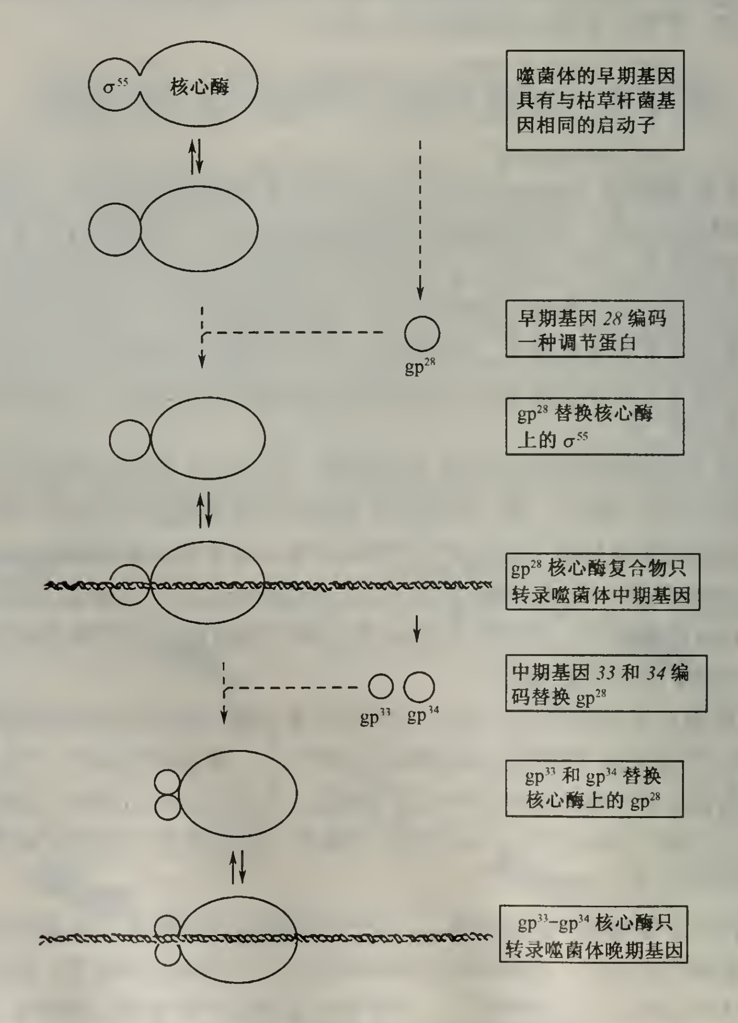


图 9-28　两种不同的 σ 因子控制 PSO1 的转录

早期基因 *28* 的突变株中的中期基因不被转录。基因 *28* 的产物是一种分子质量为 26 000Da 的蛋白质（称为 gp^{28}），它替换核心酶上的 σ 因子，这种替换使得早期基因的表达转换到中期基因的表达。替换后的新的全酶不再转录宿主的基因，而特异地转录噬菌体的中期基因。gp^{28}如何替换 σ^{55}而且宿主的 σ 因子发生了什么变化目前还不清楚，也许 gp^{28}对核心酶有更大的亲和力。

两个中期基因与下一步的转录相关联，无论是基因 *33* 或者是基因 *34* 发生突变，都能关闭晚期基因的转录。这两个基因产物分别是分子质量为 13 000Da 和 24 000Da 的蛋白质。这些蛋白质再次替换核心酶上的 gp^{28}。目前还不知道 gp^{33}和 gp^{34}是如何替换 gp^{28}（或残存的宿主 σ^{55}）的，但它们一旦和核心酶结合，就能只在晚期基因的启动子处启始转录。

σ 因子的替换有双重重要意义，亚基的每次改变，RNA 聚合酶能识别新一类基因而不再识别原先的一类基因；这些开关构成 RNA 聚合酶活性的改变。也许，实际上所有的核心酶只和某一时刻的 σ 因子形成聚合体形式，而且这一改变是不可逆的。

有些细菌在孢子形成期间利用新的 σ 因子。在营养生长期末期，由于培养基中的营养被耗尽而对数生长停止，这样就激发孢子形成。

孢子形成包含细菌生物合成活性的剧烈改变，与许多基因相关联。基础水平的调控是在转录中进行。一些营养生长期有功能的基因在孢子形成期间被关闭，但大部分继续表达。在孢子形成期末期，大约 40％的细菌 mRNA 是孢子形成特有的。

孢子形成需要的特有基因，可通过 SPO 突变被鉴别。这种突变允许正常的营养生长，但不形成孢子。这些不产生孢子的突变型根据其封闭孢子形成的不同阶段可被分为 SPO0（sporulation 不能开始）、SPOI、SPOII 等。一些是酶或孢子形成所必需的结构蛋白组成型突变，但另一些突变则被证明是调节转换生活方式有关的基因。长期以来，人们一直想弄清楚在孢子形成期间的这些变化以及这些变化的机制，但目前只知道这些改变发生在早期，且是酶的专一性改变的结果。

新形式的 RNA 聚合酶在孢子形成期变得有活性，它们含有营养期细胞同样的核心酶但含有与 σ^{55}不同的 σ 因子。这些变化如图 9-29 所示。

在孢子形成开始，σ^{37}代替 σ^{55}，在 σ^{37}的指引下，RNA 聚合酶转录孢子形成的第一套基因而停止转录营养期早期已开始转录的基因。

另一个 σ 因子 σ^{32}存在于营养细胞里，在孢子形成早期变得有活性，它再次指导 RNA 聚合酶从特定的启动子转录。

在孢子形成开始后大约 4h 左右在细胞中出现另一种含 σ^{29}的新的 RNA 聚合酶，它允许转录另一套基因。σ^{29}存在于营养细胞中，它可能是在 σ^{37}指导下形成的孢子形成基因产物之一，然而 σ^{29}如何能替换 σ^{37}或 σ^{55}仍不可思议。

σ^{28}是和营养细胞中的核心酶联系在一起的一个因子，它与核心酶组成 RNA

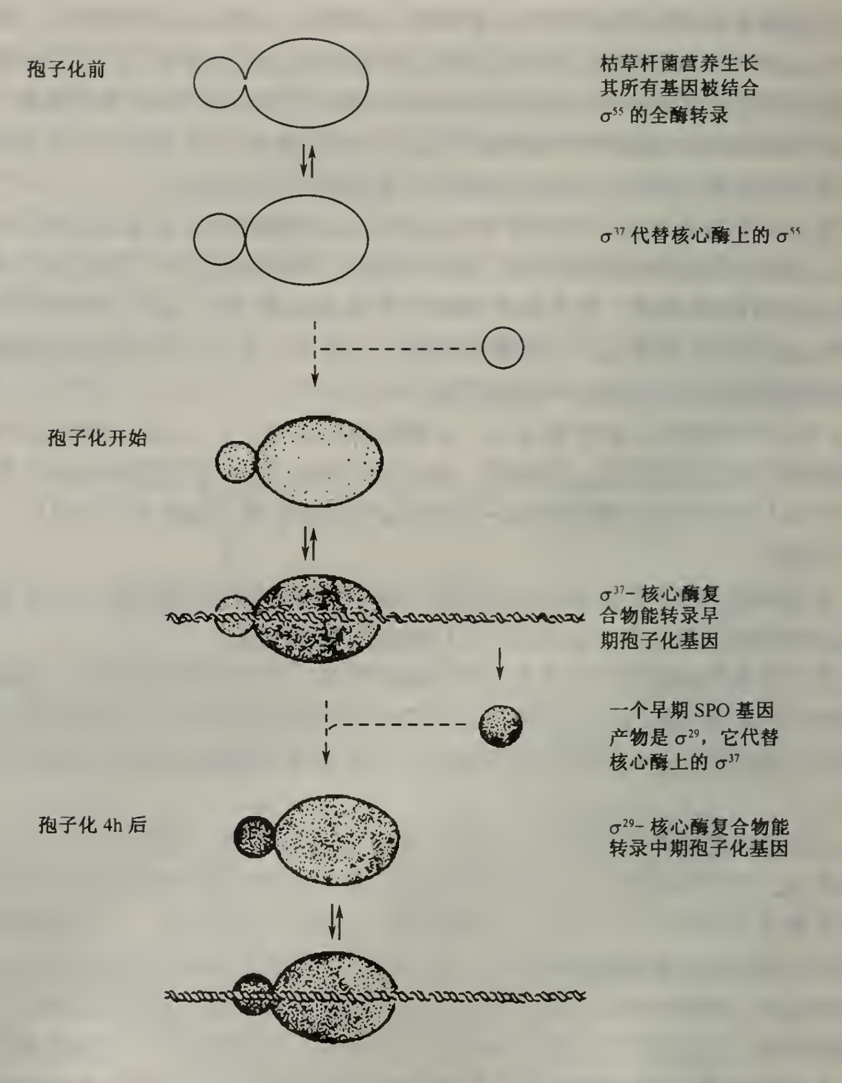


图 9-29　通过改变 σ 因子控制 RNA 聚合酶的启始专一性

聚合酶，这种形式只占全部酶活性的很少部分，当孢子形成开始时，这种形式的酶的活动就停止了。在某些 SPO0 突变株的营养细胞中，这种因子不被转录。σ^{28} 可能是某些信号系统的一部分。有的基因产物能对环境中的营养缺乏作出反应并引起孢子形成过程开始，σ^{28} 与上述的基因表达有关。因此，孢子形成过程可能受一种模式的控制，在这种模式中，σ 因子相继被激活，每种 σ 因子指导一套基因产物的合成。当一种 σ 因子变成有活性的时候，老的 σ 因子就被它所代替，这样，σ 因子的转换可以使基因开启和关闭。每种 σ 因子安排在何时出现，取决于这种 σ 因子作用的那套靶基因何时表达。这种 σ 因子的数量多少也与它的靶基因

的表达水平有关。

什么因素使得不同的 σ 因子识别不同类型的启动子呢？具有 σ^{55} 的酶识别像 *E. coli* 中那样的具有同样的 -35 区和 -10 区的启动子，而这种一致性序列在其他的 σ 因子识别的启动子中没有发现。也许每套启动子有它自己特征的一致性。目前还没有充分地研究每个 σ 因子的靶启动子的特征来归纳出所有的一致序列。一些可能的保守顺序列入表 9-3。

表 9-3　枯草杆菌中每个 σ 因子可能作用的位点

Sigma 因子		-35 区	-10 区
σ^{55}	营养体	TTGACA	TATAAT
σ^{37}	孢子形成中利用	AGGNTTT	GGNATTGNT
σ^{32}	孢子形成中利用	AAATC	TANTGTTNTA
σ^{29}	孢子形成中合成	TTNAAA	CATATT
σ^{28}	孢子形成中不利用	CTNAAA	CCGATAT
gp^{28}	SPOⅠ中期表达	AGGAGA	TTTNTTT
$gp^{33\sim34}$	SPOⅠ晚期表达	CGTTAGA	GATATT

每一种酶识别的启动子都有一个明显特征：它们的大小和距离转录起始位点的相对位置是相同的。可以看出它们保守顺序只围绕 -35 和 -10 区这个中心。-10 区的一致顺序通常富含 AT，-35 区的一致顺序也表现出这种倾向。目前还不知道不同的 σ 因子指导识别不同启动子的能力到底有多大普遍性，这一点在大肠杆菌中是罕见的，但在枯草芽孢杆菌中则具有这种特征。关于通过 σ 因子替换改变转录专一性在一些情况中已被证明，但在有些情况下又显得较复杂。所以目前还不能下结论说 σ 因子可单独起选择启动子的作用。

第七节　转录后的调控

一、翻译水平上的调控

翻译水平上的调控机制大致可分两种类型。一种为固定装置式调控（不受环境影响），即这种调控机制早已包含在 DNA 顺序内，如稀有密码子的调控和重叠基因的调控；另一种称为非固定装置式的调控（受环境影响），这种调控并不直接与 DNA 顺序有关，如，真核 mRNA 的 poly（A）长短对翻译的调控，血红素对球蛋白合成的调控等。

1. 稀有密码子和翻译调控

大肠杆菌 DNA 复制时冈崎片段之前的引物 RNA 是由 *dna G* 基因编码的引

物酶催化合成的，这种酶在细胞内所需要的量不多，如果细胞内引物酶太多还将导致细胞死亡。在大肠杆菌基因组上 *dnaG*、*rpoD*（编码 RNA 聚合酶 σ 亚基）和 *rpsU*（编码核糖体 30S 上的 S21 蛋白）同属一个操纵子，但奇怪的是三个基因产物的量却大不相同，*dnaG* 的产物仅 50 拷贝，*rpoD* 为 2800 拷贝，*rpsU* 则达 40 000 拷贝之多，细胞通过翻译的调控，解决了这个问题。

考察 dnaG 的密码发现其中含有不少稀有密码子，也就是说这些密码子在其他基因中利用频率很低，而在 dnaG 中却很高，分别计算出 *E.coli* 中 25 种非调节蛋白和 dnaG、rpoD 的密码中 64 种密码子的利用频率，可以看出 dna G 与其他两类有明显不同。例如三种 Ile 密码子中在 25 种非调节蛋白中共有 405 个，其中 AUU 占 37%、AUC 占 62%、AUA 占 1%；三种 Ile 密码子在 dna G 中共 22 个，其中 AUU 占 36%、AUC 占 32%、AUA 占 32%；在 rpoD 中则分别为 26%、74%、0%。由此可见，在 25 种非调节蛋白和 rpoD 中对 AUA 的利用频率极低，但在 dnaG 中则恰恰相反，对稀有密码 AUA 的利用频率相当高。此外，UCG（Ser)、CCU（Pro)、CCC（Pro)、ACG（Thr)、CAA（Gln)、AAT（Asn）和 AGG（Arg）等七个密码子也有一定程度的显著区别。

其他有关调控蛋白如 lacI、araC、trpR 等基因产物在细胞内的量也是很少的，这些基因中密码子的利用频率和 dnaG 相同，而明显不同于非调节蛋白。

为什么稀有密码子利用频率高的基因中蛋白质合成量低，可能是由于细胞内与稀有密码对应的 tRNA 少，这样就延长了核糖体在 mRNA 上的时间，从而降低了翻译速度，不过是否如此，还有待进一步的实验证明。

2. 重叠基因与翻译调控

重叠基因最早是在 *E.coli* 噬菌体 *ΦX*174 中发现，例如 *B* 基因包含在 *A* 基因内，*E* 基因包含在 *D* 基因内，用不同的阅读方式得到不同的蛋白质，当时认为重叠基因的生物学意义是它可以包含更多的遗传信息。后来发现丝状的 RNA 噬菌体、线粒体 DNA、插入序列，甚至在细菌染色体上也有重叠基因存在，说明重叠基因可能对基因表达的调节起着重要作用。色氨酸操纵子中的 *trpE* 基因和 *trpD* 基因之间的翻译偶联现象可以说明这个问题。

色氨酸操纵子由五个基因（*trpE*、*trpD*、*trpC*、*trpB*、*trpA*）组成，在正常情况下操纵子中五个基因产物是等摩尔的，但发现 *trpE* 突变后，其邻近的 *trpD* 的产量比下游的 *trpB*、*trpA* 的产量要低，这种差别不能用转录水平的极性来解释，遗传学实验证明这种极性效应是在翻译水平上的，当 *trpE* 能翻译时，*trpD* 也能翻译，这种依赖性便称为翻译偶联。

比较 *trpE* 和 *trpD* 以及 *trpC* 和 *trpB* 两对基因中核苷酸顺序为翻译偶联现象提供了线索，*trpE* 基因的终止密码和 *trpD* 的起始密码之间只相隔一个核苷酸，是重叠的。

而在 *trpC* 和 *trpB* 之间相隔 11 个核苷酸，由于 *trpE* 与 *trpD* 的终止密码与起始密码重叠，*trpE* 翻译终止时的核糖体立即处在起始环境中，这种重叠的密码便是保证同一核糖体对两个连续基因进行翻译偶联的机制。但对相隔几个核苷酸的 *trpC* 和 *trpB* 则不存在这种翻译偶联。

为什么说翻译偶联是一种调控机制呢？我们知道在细菌细胞内 *trpE* 和 *trpD* 的产物分别以等摩尔结合成具有功能的四聚体，而 *trpC* 和 *trpB* 则不形成复合物。*trpB* 和 *trpA* 的产物也是以等摩尔结合，虽然现在还没有发现翻译偶联现象，但这两个基因的核苷酸顺序也是重叠的。

色氨酸操纵子中 *trpE* 和 *trpD* 之间的翻译偶联现象并不是独一无二的例子。*E.coli* 的半乳糖操纵子中 *galT* 和 *galK* 在结构上存在着另一种重叠现象，*galT* 的终止密码子并不与 *galK* 的起始密码重叠，它们之间相隔三个核苷酸，但在终止密码子之间却有 *galK* 基因的核糖体结合顺序（SD 顺序）。

虽然这一情况与 *trpE-trpD* 的翻译偶联并不完全相似，但也并不能排除翻译偶联的存在。因为核糖体结合在 mRNA 上可覆盖 20 个核苷酸，包括了 SD 顺序和后面 *galK* 基因的起始密码，所以当 *galT* 终止时，核糖体还没有脱落就直接与 SD 结合开始 *galK* 的翻译。应用基因操作技术的实验证明 *galT-galK* 的确存在翻译偶联。

3. 翻译的阻遏

蛋白质作为阻遏物或激活物对转录进行调控的例子已经屡见不鲜，那么蛋白质是否也能对翻译起类似的调控作用呢？在大肠杆菌的 RNA 噬菌体 Qβ 中确实发现这种现象。Qβ 噬菌体的基因组包含有三个基因，从 5′到 3′方向依次是与噬菌体组装和吸附有关的成熟蛋白基因 *A*，外壳蛋白基因和 RNA 复制酶基因。当噬菌体感染细菌 RNA 进入细胞后，这条称为（+）链的亲本 RNA 立即作为信使合成复制酶，它与宿主中的几个亚基结合后便能起复制功能，但是这里发生了矛盾，因为在进行翻译时，Qβ（+）RNA 链上已经占据了不少核糖体，它们从 5′向 3′方向进行翻译，这些阻碍了复制酶催化的从 3′向 5′方向进行的（−）链合成，克服这个矛盾的办法便是由 Qβ 复制酶作为翻译阻遏物进行调节。

离体实验证明纯化的复制酶可以和外壳蛋白的翻译起始区结合，抑制蛋白质的合成。因此设想核糖体不能与起始区结合，但已经起始的翻译仍能继续下去，直到翻译完毕，核糖体脱下，于是复制酶便可以和（+）RNA 的 3′结合而进行复制。

这里复制酶既能和外壳蛋白的翻译起始区结合又能和（+）RNA 的 3′端结合，那么这两个部位是否有同源性呢？通过计算机算出这两个区的最稳定的二级结构，表明它们的二级结构中茎环结构相似，而且环上都有 CUUU 和 UAAA 顺序，这种相似可能和复制酶作为翻译阻遏物有关。

另一类 RNA 噬菌体 R17 和 MS2 中也存在类似于 Qβ 的翻译阻遏现象。离体实验证明外壳蛋白可以和噬菌体的 RNA 合成酶翻译起始区结合从而阻遏 RNA 合成酶的翻译。

4. mRNA 的高级结构对翻译的调控

以 RNA 噬菌体 f2 的纯化的 RNA 作为信使在大肠杆菌的无细胞系统中进行蛋白质合成时，大部分合成外壳蛋白，合成的 RNA 合成酶只占外壳蛋白的 1/3。其他 RNA 噬菌体如 MS2 也有同样现象，用同位素标记可以考察 RNA 噬菌体几种蛋白质的起译程度，结果表明外壳蛋白的起译程度比合成酶要大三倍。

另外还发现 RNA 噬菌体 f2 外壳蛋白基因的琥珀突变，除影响外壳蛋白本身外还影响到合成酶的起译。但如果琥珀突变不是发生在接近起始翻译区内而是发生在第 50、第 54 或第 70 个氨基酸处（共 129 个氨基酸），那么就不影响 RNA 合成酶的合成，这一现象说明合成酶的起译只与外壳蛋白的起译有关，与外壳蛋白 C 端的合成无关。也就是说在大肠杆菌的无细胞系统中，在完整的 f2RNA 分子上核糖体可以直接起译外壳蛋白，但不能立即起译合成酶，一定要在外壳蛋白合成到一定程度时才开始合成合成酶。我们知道在组装噬菌体时需要很多外壳蛋白，这和离体系统中观察到的现象是一致的。它通过什么机制达到这种调控呢？有一种假设是比较合理的，认为合成酶的起译部位被 RNA 的高级结构所掩盖，当外壳蛋白翻译出前面部分时 RNA 的立体结构解开，合成酶才有可能起译。用甲醛处理 RNA 可以增加合成酶的起译，说明高级结构对基因调控的可能性。

5. MIC-RNA 和翻译调控（mRNA-干扰互补 RNA）

大肠杆菌中通过转录水平上的基因调控往往是由蛋白质与 DNA 的相互作用从而达到阻遏或激活转录作用，近年来发现与渗透性有关的蛋白质以及一种转座酶是通过 RNA-RNA 相互作用对翻译进行调控的。大肠杆菌中与渗透性有关的外膜孔蛋白的调节子是由 *ompB*、*ompC* 和 *ompF* 三个相互不连锁的座位所组成。其中 *ompC* 和 *ompF* 是结构基因，编码两种结构蛋白，*ompM* 座位对 *ompC* 和 *ompF* 起调控作用。*ompB* 座位实际上是由 *ompR* 和 *envZ* 两基因组成。*envZ* 编码的蛋白设想是外膜上的一种受体蛋白，能感受培养基中渗透压的变化，并将此信息传递给细胞质内的 ompR 蛋白，通过 ompR 蛋白对 *ompC* 和 *ompF* 基因进行正控制。这里要介绍的是当培养基中渗透压变化时 ompF 蛋白产量下降，同时 ompC 蛋白的量上升，从而保持 ompC 和 ompF 两种蛋白总量恒定，即 *ompC* 基因上游 DNA 顺序转录一段 RNA 与 *ompF* 基因的 mRNA 互补，从而降低了 *ompF* 的翻译。

通过对 *ompC* 和 *ompF* 基因的核苷酸顺序分析，发现在离 *ompC* 基因不远的上游存在另外一个转录单位（存在反方向的 −10 和 −35 区）。当培养基中渗透压

高时，*ompC* 可双向转录，除转录 *ompC* 基因外，还可以相反于 *ompC* 基因的转录方向转录 174 个核苷酸的 RNA。这段转录物与 *ompF* 的 mRNA 5′端有较强的同源性，设想在细胞内通过分子杂交而抑制 *ompF* 的 mRNA 的翻译，这段 174 个核苷酸顺序称为 mic RNA（mRNA-interfering complementary RNA），由于它干扰的是 *ompF* 基因的 mRNA 的翻译，所以编码 *mic* RNA 的基因称为 *micF*，类似于 RNA-RNA 相互作用进行调控的另一个例子是转座子 Tn10 中 IS10 的表达，遗传学实验表明载有 IS10R（右边的 IS10）的向外 180bp 的多拷贝质粒可以抑制细菌染色体上 Tn10 的转座，这种抑制作用是由于转录出的 RNA 与转座酶基因的 mRNA 互补，从而干扰了翻译的进行。

6. 魔斑核苷酸的调控作用

核糖体蛋白（r-蛋白）的反馈阻遏保证了几种 r-蛋白合成的协同进行，这里主要讲如何通过魔斑核苷酸的调节使细胞内蛋白合成的总速率和蛋白合成机器中的成分 r-RNA 的合成速率之间达到协调（就是当不进行蛋白合成时，也就没有必要合成 r-RNA）。发生这种调节作用的最早信号是不载有氨基酸的 tRNA，因此这种调节也可以归纳为翻译水平上的调节。

早在 1952 年 Sands 等报道大肠杆菌缺陷型（$trp^- his^-$）在缺少任何一种所需要的氨基酸的培养基上培养时，不但蛋白质合成速率立即下降，而且 RNA 合成速率也下降，因为色氨酸和组氨酸不是 RNA 的前体，所以说 RNA 合成速率下降是由于蛋白质合成受阻后的次级反应。几年以后又发现大肠杆菌的另一个突变型，和上述情况恰恰相反，在缺少氨基酸供应时蛋白质合成虽然停止，但 RNA 合成速率并不下降，于是把前一种称为严紧控制，基因型为 rel^+；后一种现象称为松弛控制，基因型为 rel^-。

10 多年后发现 rel^+ 和 rel^- 菌株除上述生理现象不同外，当缺乏氨基酸时 rel^+ 菌株能合成鸟苷四磷酸（ppGpp）和鸟苷五磷酸（pppGpp），rel^- 菌株则不合成。因为这两种化合物是在层析谱上检出的斑点，所以当时称为魔斑（magic spot）。在旺盛生长的细胞中约有 65%～90% tRNA 是载有氨基酸的。当氨基酸缺乏时，不负载氨基酸的 tRNA 增多，这种不负载氨基酸的 tRNA 仍能和核糖体的 A 位结合，核糖体上不载氨基酸的 tRNA 是细胞产生严紧控制的讯号。在正常的蛋白质合成过程中，aa-tRNA 转移到正在延伸的多肽上时需要 GTP，也许是由于这一反应不进行了，GTP 便可用作合成魔斑的前体：

$$GTP \rightarrow pppGpp \rightarrow ppGpp$$

参与这个反应的除 *relA* 基因所编码的 ATP:GTP 3′焦磷酸转移酶外，还需要翻译起始因子 EF-Tu 和 EF-G 等。

ppGpp 的效应是多种多样的：如对许多酶（如腺苷酸琥珀酸合成酶、嘌呤磷酸核糖转移酶等）和翻译起始因子的抑制或激活；对 rRNA、tRNA 转录的抑制；

对 *lac*、*his*、*trp* 等操纵子 mRNA 的激活。虽然 ppGpp 是多效性的，但有人认为 ppGpp 对 RNA 聚合酶启动子专一性的影响可能是严紧控制的关键。

真核生物细胞培养时从高营养培养基转入低营养培养基后也观察到与原核生物中相似的多种效应，如稳定的 RNA 合成受到抑制，葡萄糖运输、核酸前体的吸取都受到抑制，同时蛋白质分解被激活，但真核生物中尚未发现 ppGpp。

有人将 ppGpp 和 cAMP 这类物质称为警报素（alarmone），当细胞缺乏氨基酸时产生 ppGpp，于是细胞对它可以做出种种反应以求缓和紧急的情况：抑制核糖体或其他大分子的合成；活化某些氨基酸操纵子的转录；活化蛋白水解酶；抑制与氨基酸无关的转运系统等，从而节约能量和原料或开发能量或原料的来源。

除了上述几种方式的翻译水平调控外，还值得一提的是某些多顺反子mRNA几个基因的差别性翻译。例如 *lac* 操纵子转录出一条多顺反子 mRNA 在充分诱导的细胞里 β-半乳糖苷酶、透性酶和转乙酰酶活力的比例是 1∶0.5∶0.2，这很可能是细胞对这三种酶需要量的反映，在翻译过程中 *lac* mRNA 会经常脱离核糖体，对于下游的一个起始密码子（AUG）需要重新起始，于是从 5′端到 3′端的几种蛋白质的合成形成了梯度，还可能由于另外的原因，即 *lac* mRNA 的 *a* 基因部分更容易受到内切核酸酶的降解，因此 *Z* 基因蛋白的合成就比 *a* 为多。这种梯度现象是亿万年进化的结果。

二、翻译后调控

翻译后调控主要指两方面的内容，一方面是指无论是原核生物还是真核生物的蛋白质在细胞内都有一定的分布部位，大部分存在于细胞质内，其余则分布在细胞的其他组分上，像大肠杆菌这样简单的细胞，蛋白质分布在细胞质、内膜、外膜和周质等四个区域内。细胞质内约占 75%；内膜上的蛋白总量虽少但种类颇多，包括与物质运输、电子传递以及与脂质合成有关的蛋白；外膜上蛋白质种类不多，但总量比内膜多，不少是作为受体的蛋白。对于真核细胞来说蛋白分布区域还包括一些细胞器或亚细胞结构的膜，如线粒体、叶绿体以及粗糙内质网等。

可是我们知道蛋白质的合成机器在细胞质内，这些蛋白质是怎样进入到膜或周质中去呢？20 世纪 70 年代提出信号假设对这个问题做出了解答，认为分泌性蛋白质 N 端的一段称为信号肽的几十个氨基酸具有导航的功能，它能引导细胞质内正在合成中的蛋白质进入膜或周质中，这段信号肽则在蛋白质合成过程中途被水解去除。

通过对癌细胞的免疫球蛋白 IgG 的研究发现，IgG 轻链的 mRNA 全部存在于和膜结合的核糖体上，并且证明离体系统中转录时的翻译产物比真正轻链大些，提出了信号肽假设，要点是：①只有那些为着膜蛋白编码的 mRNA5′端有额

外的一段密码子顺序；②当这段顺序被译出来以后，核糖体就会附着到膜上，在膜上形成通道的小孔，从而使新生肽通过膜，如果没有这段肽，核糖体不会附着到膜上，这段肽称为信号肽；③信号肽长度约为 15～30 个氨基酸。其中以疏水性氨基酸占多数；④当信号肽进入膜后肽链继续延伸，在尚未完成之前信号肽先被水解去，膜上可能存在着这种具有专一性的蛋白水解酶；⑤当多肽全部合成以后便从核糖体释放到膜上或周质中，于是核糖体又从膜上游离出来。

目前对信号肽的结构和功能已研究得比较清楚，并且已将信号顺序用于获得更多分泌性产物的基因工程中。

翻译调控的另一方面主要是指翻译后的蛋白质修饰及构象形成。如从mRNA直接翻译出的多肽往往无生物学活性，经切除一部分后具有活性；有些蛋白质翻译后需经甲基化、糖基化之后才具活性，这种修饰作用能使 20 种氨基酸变得多种多样，修饰后可达 100 多种；有些酶（如激酶）需要磷酸化后才有活性等。翻译产生的多肽链还必须正确折叠形成正确的构象才具有生物学功能。最初人们由于被核酸酶热变性失活冷却后又可恢复正确构象而重新具有活性所误导，认为蛋白质可自动形成正确构象。近年来的研究发现表明，蛋白质的构象形成至少与环境及一类负责帮助蛋白质形成正确构象的被称为“蛋白质伴娘”的蛋白有关。蛋白质的氨基酸排列顺序完全相同但可形成完全不同的构象而具不同的性质与功能，朊病毒（prion）蛋白就是一个例子，因目前关于蛋白质形成构象的机制和相关因子认识太少，所以不能在此介绍和讨论。

第十章 真核生物基因组及其基因表达调控

我们目前了解的关于遗传信息的储存、复制与表达的知识大部分仍是来源于对原核细胞和病毒的研究。但近年来越来越多的人从事于真核生物的基因组、基因结构及表达调控方面的研究。这方面的某些研究虽然不像对原核生物的研究那样深入，那样明确。但是在真核基因组的特点和真核基因的结构方面的研究都取得了重大成果，同时对真核基因表达调控方面的研究也取得了一些重要的进展。

第一节 真核生物基因组

和原核生物相比，真核生物基因组有以下特点：①基因组大，具有多个复制起点；②一个基因组包括若干个染色体，一般不呈环状；③整个染色体 DNA 分子都和蛋白质稳定地结合；④有大的重复序列；⑤功能上密切相关的基因的集中程度不如原核生物，还没有发现像原核生物中那样的操纵子存在。

一、重复序列

对真核生物基因组的最早研究是从复性动力学和密度梯度离心开始的。在最早的对真核生物 DNA 和原核生物 DNA 复性动力学研究中发现，真核 DNA 复性与原核 DNA 很不相同，在真核生物中 DNA 复性时，其复性曲线除了与原核 DNA 部分相似外，还有一部分是复性很快的部分，复性最快的这部分就是重复序列。通过 DNA 的复性动力学曲线可粗略地估计基因组的大小以及这个基因组中单拷贝基因和重复 DNA 的量，因为在复性过程中，基因组越大复性越慢。另外重复 DNA 比单拷贝基因复性也要快。所以根据复性动力学实验及结果可以粗略估计基因组的大小和重复程度。

从重复的程度来说，真核生物的 DNA 序列大致上可以分为：单拷贝序列、少量重复序列、中度重复序列和高度重复序列。

（1）*单拷贝序列*

这部分是复性曲线中最后复性的部分，占基因组的 40%～70%。

（2）*少量重复序列*

每一种的重复次数大约 1～10 次，例如编码组蛋白和 tRNA 的基因。某些少量重复基因的核苷酸顺序并不完全相同。这里包括两种情况：①几个基因都是有功能的，例如血红蛋白 β 链有胚胎型的（2～8 周前的胚胎）、胎儿型（指 8～41

周的胎儿）和成人型的，这些蛋白质都只有几个氨基酸的差别；②有些基因有功能，有些则没有功能，如假基因。

(3) 中度重复序列

这些序列一般是分散的，每一种序列的重复次数不等，大约在 10 到几百个拷贝。它们的排列有方向一致的，也有方向相反的，许多中度重复序列和单拷贝序列相间排列，也有中间没有单拷贝序列的。

中度重复序列大致可分为两类：短周期散置重复序列和长周期散置重复序列。人、爪蟾、海胆的中度重复序列以前者为主，果蝇的中度重复序列以后者为主。

短周期序列每序列长约 100～300bp，长周期序列的每序列大约 5000bp。长短周期序列均有若干种，但它们往往有共同的结构特征。例如 5000bp 的重复序列的两端各有长度约 255～400bp 的顺向重复序列。这种排列方式在酵母中也有发现。这种现象和下面要讲到的鸟类和哺乳动物染色体上整合的反转病毒十分相似，也和细菌的转座因子两侧的顺向重复序列相似，这些重复序列可能类似于转座因子的成分。

在果蝇中曾发现某些中度重复序列在黑腹果蝇和它的近缘种间在数目和位置上都有不同，而且在黑腹果蝇的一些品系间也有不同。这些情况也说明这些中度重复序列中间可能包括转座子。

(4) 高度重复序列

这些序列一般不分散，大部分集中在异染色质中，特别是在中心粒和端粒的附近。高度重复序列中常有一些和大部分 DNA 具有不同的 G+C 含量的所谓卫星 DNA。许多真核生物都有卫星 DNA（satellite DNA），它约占总 DNA 的 1%～30%，一些卫星 DNA 的核苷酸顺序已经测定。例如牛的卫星 DNA 的绝大部分是 ATATAT……。一种果蝇（drosophila virils）有三个卫星 DNA 区带，它们的核苷酸顺序分别是：5′-ACAAATT-3′；5′-ACAAACT-3′；5′-ATAAACT-3′。这些顺序重复约 10^7 次。

许多真核生物的核内 DNA 在 CsCl 或 Cs_2SO_4 中密度梯度离心时，除了形成一个主带外，往往在主带的轻侧或重侧还有一个或几个小带，这就是卫星带。对这部分 DNA 进行分析发现是重复顺序，这部分 DNA 称为卫星 DNA。图 10-1 是大肠杆菌和蟹类核内 DNA 在 CsCl 密度梯度离心时出现的密度峰，大肠杆菌中只有一个峰，而蟹类则除了密度为 1.705 的一个峰外，还有一个密度为 1.683 的峰。

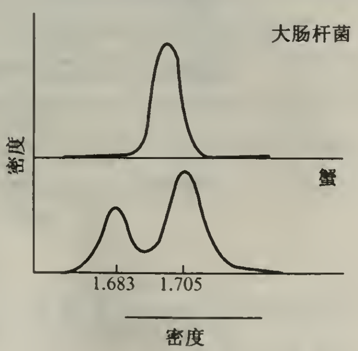


图 10-1 蟹类核内 DNA 和 *E.coli* DNA 在 CsCl 梯度离心中的密度峰

用复性实验证明卫星DNA全部是高度重复顺序，而高度重复顺序则有大部分（不是全部）包含在卫星DNA中。把小鼠卫星DNA通过RNA聚合酶在体外转录成有放射性标记的互补RNA，以此RNA与体外培养的小鼠细胞做原位杂交，放射自显影后证明卫星DNA分布在20对染色体中所有19对常染色体和X染色体的着丝点（centromere）附近，即细胞学中称为异染色质（heterochromatin）部分。Y染色体上没有卫星DNA，但Y染色体上很多部分是异染色质。这说明，卫星DNA在异染色质部分，但异染色质上的DNA并非全部是卫星DNA，异染色质上还有非卫星DNA和单一顺序。

二、基因家族

真核生物中虽然很少有真正的操纵子，可是结构和功能上相关的基因常聚集在一起形成所谓基因家族（gene familie），虽然同一家族的成员也可以分居各处。属于一个基因家族的各种基因可能都是单拷贝的，也可能是多拷贝的。

根据基因家族的复杂程度，可以把它们分为下列几种类型：简单的多基因家族、复杂的多基因家族、不同场合表达的复杂的多基因家族（图10-2）。

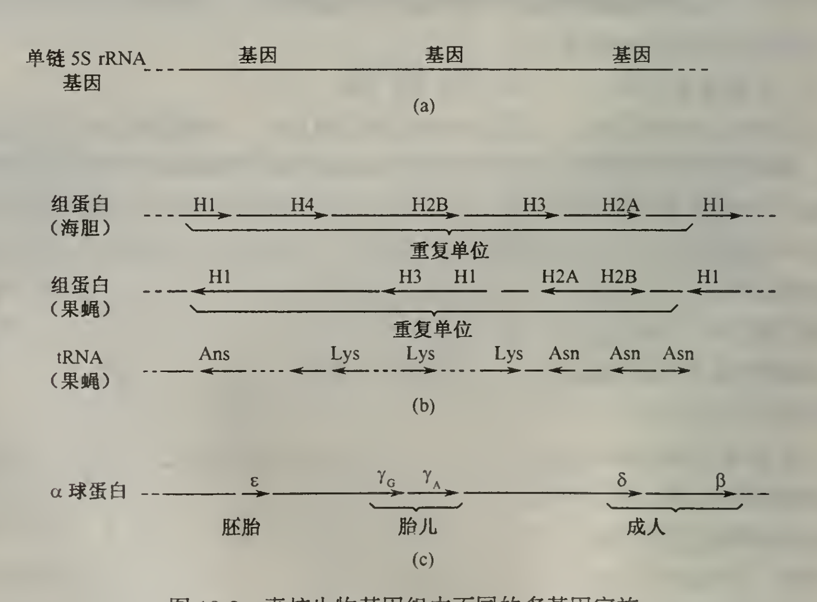


图10-2　真核生物基因组中不同的多基因家族

(a) 简单的多基因家族；(b) 复杂的多基因家族；(c) 发育控制的复杂多基因家族。

一些生物的5S RNA基因构成简单的基因家族。它们的结构相同，基因与基因之间是中度重复顺序。

复杂的多基因家族的各个成员并不都相同。海胆和果蝇的组蛋白基因以及果

蝇的 tRNA 基因就属于这类基因家族。五个组蛋白基因作为一个单位重复达千次，每组基因单独地按一定方向转录。某些 tRNA 基因集中在一起，另一些则集中在其他位置。遗传学分析证明玉米有三类醇溶性蛋白，其基因数目分别是 24、52 和 2，它们分别处在三个染色体上。一些基因是分散的，而另一些则是集中在一起的。

人的珠蛋白基因家族是属于不同场合表达的复杂多基因家族，包括胚胎时期表达的珠蛋白基因，胎儿时期表达的 γ 珠蛋白基因和成体中表达的 δ 和 β 珠蛋白基因。

以上这些基因家族中的成员都是有功能的，尽管有些基因可能只在某个发育阶段表达。而在有一些基因家族中则包括一些没有功能的成员，这些便是假基因。假基因是和一般编码的基因在结构上相似，可是并不表达的基因。这些基因可以认为和表达基因属于同一个家族。

人们早就设想新的基因来源于已有基因的重复，重复基因的命运可以是以下 3 种：①具有和原有基因几乎完全相同的结构和功能，如珠蛋白基因 *α-ad1* 和 *α-ad2*。②重复基因发生较多的结构上的变化，可是仍然执行几乎相同的功能。例如血红蛋白的 α 链和 β 链基因。③重复基因的结构和功能都发生了较大的变化，如编码肌红蛋白和编码血红蛋白的基因。

假基因虽然也可以认为是原有基因的重复，可是和其他重复基因有显著的区别，它们和有功能的基因相比除了一些一般性的变化以外往往还有一些特殊变化：①缺少内含子；②在假基因的末端连接着许多 dAMP；③两端有顺向重复序列。

根据这些结构上的特点，一般认为假基因似乎不是正常基因的直接重复，而是正常基因的 mRNA 经剪接以后再经过反转录酶的作用合成相应的 DNA，然后再整合到染色体上的结果。而且整合方式也不同于一般的同源重组，因为在它的两侧出现顺向重复序列。在这一点上假基因和下面要讨论的反转病毒相类似。

对人的 DNA 进行琼脂糖电泳分析的结果说明中度重复序列中除了一些长短不等的重复序列以外，总的 DNA 中大约有 3%的重复序列长度相等，都是大约 300bp。一个基因组中拷贝数大约是 $3\times10^5\sim5\times10^5$。在这部分重复序列中 60%可以为限制酶 *Alu* 所识别，说明这是一类长度、性质相似的重复序列，因此称之为 Alu 家族。

随机取人的 DNA 克隆 100 个，每个克隆中的 DNA 长度是 $1.5\times10^4\sim2\times10^4$bp，用 Alu 家族成员 DNA 作为探针去检测，发现 95%的克隆能和它进行杂交。说明这些片段中部包含着 Alu 家族成员。同样也说明了 Alu 家族的普遍性，它是中度重复序列中最普遍的一种。

Alu 家族不只在人基因中发现，非洲绿猴、小鼠、中国仓鼠等哺乳动物中都有。比较这些动物的 Alu 家族成员的核苷酸顺序，可以看到它们的大部分顺序是

相同的，特别是3′端都有富含dAMP的序列。

这些Alu成员在它们的两端都具有7～20bp的顺向重复序列，各个Alu成员的两端的顺向重复序列都不相同。以下列出两个属于人Alu成员的核苷酸顺序：

GTTTAGATAAG……Alu……A_{25}GTTTAGATAAA

AAAGAAATG……Alu……$A_{14}GA_3GA_4GA_5GA_6GA_5$

AAATAAATGGAlu序列两端的顺向重复顺序和大肠杆菌中已发现的转座子两侧的顺序相类似。有报道说同一个体的不同组织中Alu序列数不相等，这一现象和酵母菌中的转座因子Ty-l的情况相似。

除了Alu家族以外哺乳动物染色体上还有其他较小的家族。

三、反转病毒和癌基因

反转病毒是一类单链线状RNA病毒，它们通过反转录酶的作用合成DNA中间体，DNA中间体经细胞质进入细胞核，然后整合到染色体DNA上而成为原病毒（provirus）。

反转病毒的基因组包括两个相同的线状单链RNA，各长5～10kb（图10-3），感染宿主后通过反转录酶的作用合成一个双链DNA（图10-4），它的两端出现两个较长的重复序列（LTR），每个重复序列包括原来属于病毒基因组亚单位的两个不同的部分U3、U5和相同部分R（图10-4）。

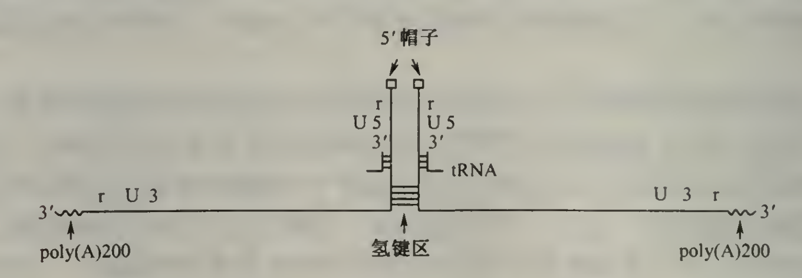


图10-3　Rous肉瘤病毒核酸的结构

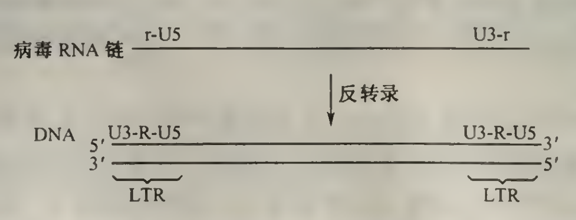


图10-4　38S病毒RNA链上的末端序列及其DNA拷贝上的长末端序列（LTR）

反转病毒RNA经反转录后形成的DNA两端存在重复序列暗示反转病毒和

转座子有类似性质。为了进一步研究病毒的整合过程，采用了基因克隆手段来分析原病毒和宿主染色体 DNA 交接处的核苷酸顺序，结果发现在小鼠乳房瘤病毒（mouse mammary tumour virus，MMTV）每个 LTR 序列的两端都有 5bp 的反向重复序列。病毒 DNA 与宿主 DNA 交接位置处宿主 DNA 上有 6bp 顺向重复。LTR 的两端的反向重复序列的存在令人想到大肠杆菌的 ISI 两端的反向重复序列，顺向重复序列的存在令人想到转座子两侧的重复序列。大肠杆菌的 Tn 插入染色体两侧出现顺向重复序列，其中一个序列是染色体上原有的序列。除了 MMTV 外，其他反转病毒插入染色体后也有这些结构上的特点，如莫氏啮齿类白血病病毒和脾脏坏血病病毒。

为了进一步说明末端重复顺序的功能，在 SNV 中进行了末端缺失所带来的对于 DNA 复制或整合的影响的研究，发现 U3 末端 12bp 和 U5 末端 8bp 中一些核苷酸的缺失使病毒不能整合，但是仍然能复制，说明它们对整合是必要的，可对于 DNA 复制并不必要。这一实验结果与转座子转座所必需的结构特征相同，两端的重复顺序为转座子转座所必需，反转病毒的 LTR 则为其整合所必需，这就给我们提出了一个问题，反转病毒除了通过染色体外复制并整合到染色体上这一过程以外，是否还可以像转座子一样直接在基因组内转移位置呢？一种内源病毒（原病毒）在一个染色体上可以有若干个，这一事实说明病毒可能通过不止一种方式出现在染色体的不同位置上。前面已经指出许多长周期散置重复序列的结构和整合的反转病毒的结构相似，所以两者之间的关系是一个耐人寻味的问题。

反转病毒有许多种类。从鸟、小鼠、大鼠到爬行动物、猫、仓鼠、猴、猪、牛、鼠和人都分离到这类病毒。尤其以鸟类和小鼠为最多。有些反转病毒具有复制和繁殖所必需的全部基因，因此能独立复制和繁殖。另一些则缺失了某些基因，因此只有在野生型作为辅助病毒存在于同一细胞时才能复制和繁殖。

通常反转病毒含有反转酶基因（*pol*）、衣壳蛋白基因（*env*）和启动区（*c*）。除了与病毒的复制和繁殖有关的基因外，许多反转病毒的基因组中还有癌基因 *src*（肉瘤病毒癌基因）和来源于宿主细胞的癌基因 *onc*。

反转病毒感染宿主后能使宿主产生肿瘤，也能使培养中的成纤维细胞发生恶性转化。将这些转化细胞接种到动物身上能引起肿瘤生长，这是因为许多反转病毒的基因组中有 *src* 或 *onc* 基因的缘故。

鸟造白细胞组织增生病毒（ALV）基因组中虽然既无 *src* 基因又无 *onc* 基因，可是经过很长潜伏期后也能引起肿瘤，这是由于它整合在染色体上的一个癌基因的近旁而使它被激活的缘故。人和动物的癌基因便是从反转病毒的研究开始的。

癌基因的发现是反转病毒研究和分子生物学技术相结合的结果。应用基因克隆、Southern blot 和分子杂交方法首先于 1976 年发现鸟类的基因组 DNA 中具有与劳氏肉瘤病毒的 *src* 基因核苷酸顺序相近似的序列，接着又发现在哺乳动物的

DNA 中也有这些序列。随着反转病毒中更多癌基因的发现，同时也证实了这些癌基因的同源序列都可以在动物的基因组中找到，为了便于区别，Cofffin 等于 1981 年提出了一个统一命名规则。把属于病毒的癌基因称为病毒癌基因（viral oncogene，*V-onc*），把属于动物细胞的癌基因称为细胞癌基因（cellular oncogene，*C-onc*），把来自某种病毒的癌基因用病毒名称的缩写表示。例如把来自成髓细胞性白血病病毒（myeloblastoma virus）的癌基因称为 *myb*，把来自成髓细胞瘤病毒（myelocytomatosis virus）的癌基因称为 *myc*，把来自肉瘤病毒（sarcoma virus）的癌基因称为 *src*，把来自大鼠瘤病毒（rat sarcoma virus）的癌基因称为 *ras*，有时两种癌基因来自同一种病毒，那么它们分别用 A、B 表示。例如来自红细胞增殖病毒（erythroblastosis virus）的两种癌基因称为 *erb-A*、*erb-B*。原癌基因（proto-oncogene）指具有致癌活性基因的无活性原型。将近 20 年来的研究结果表明正常细胞中具有无活性的原癌基因。它们可以通过种种方式被活化而成为癌基因。

到现在为止已经发现反转病毒的癌基因（*V-onc*）有十几种，每一种都有它的相应的 *C-onc*，许多 *C-onc* 具有近似的核苷酸顺序。此外还发现几个不同的癌基因编码同一种酶——酪氨酸专一的蛋白质激酶。这些事实都说明正常动物基因组中存在着癌基因家族，其中可能包括 50～100 个癌基因，这些基因分布在几条染色体上。

除了在脊椎动物以外还在果蝇和线虫中也发现有癌基因，这些癌基因不但和高等动物中的癌基因在核苷酸顺序方面显示出同源性，而且可能也在功能上相似，例如在果蝇一些组织中发现酪氨酸专一的蛋白激酶，这酶也被劳氏肉瘤病毒的 *V-src* 基因和鸟类及哺乳类的 *C-src* 基因所编码。

由于反转病毒和许多动物的基因组有相同或相似的癌基因，必然会发生这样的问题：这些基因来源于病毒还是来源于动物基因组？以下几方面的事实说明癌基因来源于动物基因组：①在各种动物中癌基因的数目和位置比较固定，可是在反转病毒中的数目和位置则不固定；②线虫和果蝇等生物中发现与高等动物癌基因同源的基因，可反转病毒不能感染这些生物；③在许多高等动物癌基因中发现有内含子，反转病毒的癌基因中则没有内含子；④动物的癌基因可以被反转病毒所转导。

癌基因是有害基因，为什么真核生物基因组中包含这么多癌基因呢？一种解释是癌基因是正常生长和发育所必需的，在正常组织中癌基因常有不同程度的表达，一些实验指出这些基因的表达有组织专一性，如某一种反转病毒只能使某一些器官生长肿瘤，而且只能使某些细胞发生转化。癌的遗传倾向性也是组织专一的，许多癌基因在各种组织中或者在不同的发育阶段的表达往往不同，这些事实暗示我们原癌基因可能是一些正常发育所必需的基因，是基因组的正常成分。

病毒引起癌变，这一事实首先在 1911 年由 Ross 所发现。他发现由鸡的肉瘤

中所提取的病毒能诱发肿瘤，因此这一病毒称 Ross virus。这一发现当时并没有受到注意。

肿瘤可以由许多因素引起，这使得肿瘤起因的研究发生困难。自从发现了病毒的致癌作用是由于带有癌基因，并且由于以后证实动物细胞中存在为数不少的原癌基因，而这些原癌基因整合到病毒中以后转变为癌基因，于是癌变的真正原因的探索便出现了新的希望。既然正常细胞中存在着原癌基因，既然这些基因整合到病毒中以后会被激活，那么这些基因在不脱离正常的基因组时会不会也被激活呢？因此原癌基因被激活的问题是一个受到十分重视的课题。

最简单的情况是原癌基因整合在反转病毒的下游，在这种情况下不被转录的原癌基因便被转录，也就是说原癌基因被激活成癌基因，这些反转病毒感染动物以后整合到动物基因组上，被激活的癌基因便通过促使染色体复制和细胞分裂而造成癌变。

鸟类白血病病毒（ALV）的基因组中并没有癌基因，可是它能够引起肿瘤生长，这是因为它整合在细胞癌基因（*C-myc*）的近旁的缘故。转录产物测定结果说明 *C-myc* 的转录在其旁整合有 ALV 的情况下增加 100 倍，类似的现象也在一些其他病毒感染中发现。

没有病毒感染的情况下原癌基因怎样被激活是一个更有实际意义的问题。人的膀胱肉瘤组织原癌基因 *C-ras* 的激活是由于发生了一个碱基对的置换（G→T），它使编码的蛋白质中发生了一个氨基酸的取代（甘氨酸变为缬氨酸）。

染色体畸变可能是另一个原因。Burkitt 氏淋巴瘤（一种 B 细胞肿瘤）的细胞中有一个基因易位，它把第八染色体上的一个原癌基因转移到第十四染色体的免疫球蛋白基因的旁边，类似现象也在小鼠的浆细胞中发现，它的十五染色体上的 *C-myc* 易位到第十二染色体上。

基因激活的结果或是导致基因产物的量的增加，或是导致产物的质的变化(如酶的活性及底物专一性改变)。癌细胞的一个最主要的特性是无控制的细胞分裂，细胞分裂以前先要进行 DNA 复制，细胞的表面结构也随着癌变而发生变化，癌基因产物怎样发生影响，这些问题还有待于进一步研究解决。

四、真核细胞中的转座因子

酵母菌是低等真核生物，它的基因组虽然在许多方面与原核不同，但其转座子却在许多方面与细菌的转座子相似，包括它的遗传学效应。它能导致染色体畸变，它的两端有重复等。转座子 Ty-1（transposion yeast）的大小为 6.5kb，有两个 δ 的顺向重复序列，其长度为 0.25kb。酵母菌基因组上的 δ 序列大约有 100 个，暗示 δ 也是转座子。

在不同品系中 Ty-1 片段数目不同。Ty-1 所在的位置也不同，这些都说明

Ty-1是可以转移位置的DNA序列。Ty-1插入基因 *His4* 造成自发突变，插入位置出现 *His4* 基因两侧有五个核苷酸对的重复。

1. 果蝇的转座因子

果蝇的转座因子P因子的存在可以为宿主带来一系列的异常现象，这些现象常出现在品系间杂交子代中，所以称为杂交退化综合征（hybrid dysgenesis syndrom），如卵巢发育不全，以及雄性个体中的重组，高突变率等。

2. 玉米中的解离因子、活化因子

20世纪40年代发现玉米中有解离因子（DS）和活化因子（AC），这些因子与转座子一样引起插入突变。

五、真核细胞中的线粒体基因组和叶绿体基因组

真核细胞中除了核基因外，还有线粒体DNA和叶绿体DNA，它们是生命活动所必需的，这一点与原核生物不同，原核生物的染色体外DNA（质粒）不是细胞生活所必需的。

线粒体和叶绿体基因组除了一部分基因编码它本身的核糖体蛋白质和tRNA以外，还有一部分基因编码某些蛋白质的一些亚基。它们和核基因编码的亚基共同组成有功能的蛋白质。如玉米等植物的二磷酸核酮糖转化酶的大亚基由叶绿体基因所编码，它的小亚基由核基因所编码。

一些事实暗示我们核基因组、线粒体基因组和叶绿体基因组之间可以发生基因交流，酵母菌的ATPase的基因属于线粒体基因组，脉孢菌的编码ATPase的基因却属于核基因组。玉米线粒体基因组和叶绿体基因组有12kb的同源序列。

酵母菌的基因 *Vasl* 是一个线粒体基因组中与线粒体核糖体装配有关的基因。用 *E.coli* 质粒pBR322克隆这一基因，制备这一基因的探针，用它与干酵母核DNA进行Southern吸印杂交，发现在核DNA中有同源序列。

根据以上事实可以认为核基因组和线粒体或叶绿体基因组间可以发生基因交流。这些事实也支持共生进化学说，认为真核生物的线粒体和叶绿体来源于远古时代的共生生物。从共生开始以后可能便不断地发生着共生生物和宿主细胞之间的基因交流。

第二节　真核基因的结构

真核细胞的重要特点是细胞分化为细胞核和细胞质，整个基因组几乎全部包括在细胞核内，基因在核内以染色体的形式存在，核膜把细胞核和细胞质隔开。

对真核基因的早期研究是搞清染色体的组成、DNA 分子在染色体中的存在状态和数目等。对染色体的化学分析的结果是：大约 1/5 是 DNA，2/3 是蛋白质，还有少量的 RNA。20 世纪 70 年代改进了 DNA 的提取方法。在提取过程中尽量避免了扯断，测得 DNA 的分子质量大小刚好等于一条染色体中的 DNA 含量。通过大量实验以后，1974 年以后一致公认，每条复制前的真核染色体只有一个连续不断的 DNA 双螺旋分子。即使在着丝点处也不中断，复制后的染色体则有两个 DNA 分子，即每条染色单体有一个 DNA 双链分子。之后便开始了对真核细胞特定基因的结构及表达的研究。

一、rRNA 基因

rRNA 是细胞中主要的转录产物之一，在高等真核生物中要求有众多为 rRNA 编码的基因拷贝。这些基因都位于细胞核内。在许多高等真核生物中，这些基因一前一后地串联排列，并且在电子显微镜下可以看到在转录中的各重复单位的情况。图 10-5 是 rDNA 被转录的电镜照片。



图 10-5　爪蛙细胞核 rDNA 基因转录的电镜照片

每个转录单位被不转录的间隔区分开。被转录的区域被 RNA 聚合酶和新生的 RNA 链包围着。新生 RNA 链和蛋白质联系在一起产生前核糖体粒子。这些粒子中含有成熟核糖体中的蛋白质以及还有一些非核糖体核蛋白，这些蛋白具有内切核酸酶活性，参与 rDNA 初级转录产物的加工。

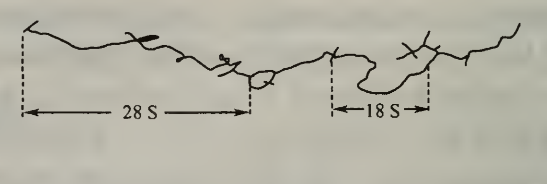
人肿瘤细胞系 Hela 细胞的 rDNA 的初级转录产物是 45S。经加工后产生 18S 和 28S（图 10-6）。

目前对几种生物的 rRNA 基因的结构及转录加工过程都研究得比较清楚。

1. 非洲爪蛙的 rRNA 基因

用原位分子杂交法也可证明 rRNA 基因在随体染色体的核仁组织区上。而 5S RNA 基因则分布在几乎所有染色体的端部。

rRNA 基因和 5S RNA 基因是最早分离成功的真核基因。这是由于它们在染色体上多次重复，并且它们的 G+C 含量比其他 DNA 高，所以很容



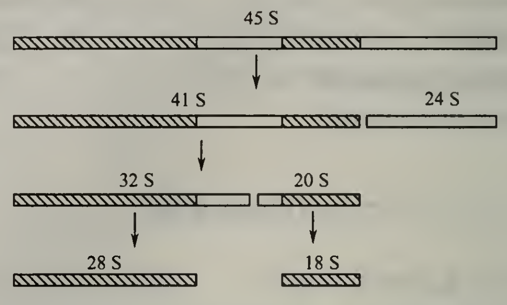


图 10-6　Hela 细胞 rRNA 的结构图

易通过 CsCl 密度梯度离心将其分离。

1966 年 M.Birnstie 等将 DNA 分离成功并进行研究后很快得出以下结论：①rRNA基因在染色体上重复多次，每个核仁组织区有约 450 个拷贝首尾相接，因此每个二倍体细胞内约 900 个拷贝。②基因和基因之间有间隔区（spacer），即不转录部分。转录子与间隔区相间排列，rDNA 应包括 rRNA 基因和间隔区。

他们用放射性标记的 18S RNA 和 28S RNA 分别与非洲爪蛙体细胞 DNA 分子杂交，发现体细胞 DNA 中有 0.04%～0.07%和 28S RNA 杂交。有 0.025%～0.04%的 DNA 与 18S RNA 杂交。已知 28S RNA 的分子质量是 1.4×10^6Da。18S RNA 的分子质量是 0.7×10^6Da，后者是前者的一半，分子杂交的量也是 18S 约为 28S 的一半，因此认为 28S DNA 和 18S DNA 的拷贝数相等。

D.D.Brown 等 1968 年用^{3}H 标记的 18S RNA 和 28S RNA 一起和 DNA 做分子杂交，结果杂交的总量为总 DNA 的 0.057%。

已知非洲爪蛙的单个基因组的 DNA 重量约为 3pg（$1pg=10^{-12}g$），假定和 RNA 杂交的 DNA 单链的重量和杂交的 RNA 量大致相等，则可计算其拷贝数。

考虑到 DNA 扯断的长度不会刚好相当于 18S RNA 和 28S RNA 的长度，因此杂交的 DNA 单链比 RNA 链略长一些，大致估计每单倍体基因组内 rRNA 基因拷贝数在 450 左右。

将 rDNA 用弱碱溶液处理，部分变性后在电镜下可以看到其部分变性图。rRNA 基因拷贝首尾相接，变性部分是 DNA 双链中 A、T 含量高的区域拆开成的单链，而 G、C 含量高的区域仍保持不变性的双链（图 10-7）。从图 10-7 上看出每一重复单位平均长 6μm，不变性段为 1μm，变性段为 5μm。已知 28S RNA 和 18S RNA 中 G+C 含量特别高，因此推测不变性区是编码 rRNA 基因部分，

变性区是间隔区。从转录初级产物为 40S（pre RNA）的分子质量大约是 2.6×10^6Da 来看，DNA 中每一转录区段应是 3.5μm，间隔区约为 2.5μm。

非洲爪蛙 rDNA 的初级转录产物为 40S。经加工后产生三种 rRNA：18S RNA、5.8S RNA 和 28S RNA。一个 rRNA 基因重复单位和它在染色体上的排列如图 10-8 所示。

图中不转录的部分为外部间隔区（SP-e），也称为不转录的间隔区（NTS），初级转录产物中在加工时切除的部分为内部间隔区（SP-i，图中斜线部分）。

变性区

非变性区

一个重复单位

图 10-7　Triturus Viridisen 游离核仁的部分变性图

2．5S RNA 基因

爪蛙的 5S RNA 基因研究得最为清楚，5S RNA 的碱基顺序已全部测定，5S RNA上的启动区顺序已测定，间隔区顺序也已知道。

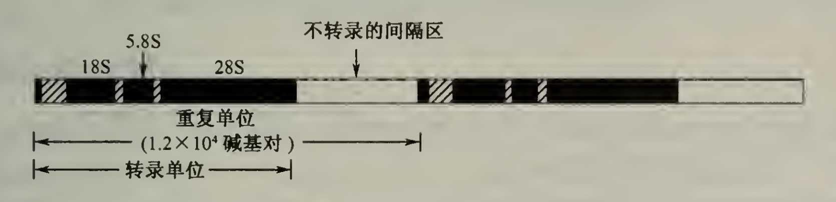


图 10-8　爪蛙 40S 前体 rRNA 基因

5S DNA 占基因组总 DNA 的 0.7%，每单倍体基因组中有 24 000 拷贝，分布在几乎所有染色体的端部，每条染色体端部有 1000～2000 拷贝。

用弱碱溶液处理 5S DNA 可得到变性和不变性部分交替出现的部分变性图（图 10-9），间隔区与转录区比例是 6∶1。5S RNA 基因在染色体上的排列也是首尾相接。5S RNA 基因每个重复单位为 720bp，非转录部分除了间隔区外还有一个假基因（图 10-10）。

爪蛙 5S 基因却有两种，卵母细胞一种，体细胞一种，相差 6 个 bp，5S RNA 121bp，假基因 101bp。假基因 101bp 中与 5S RNA 有 10 个 bp 不同，并且缺失了 5S RNA3′端的 19 个 bp，启动区有 9 个 bp 不同。图 10-11 是不同 5S RNA 基因的不同结构。有些有假基因，有些则没有假基因。

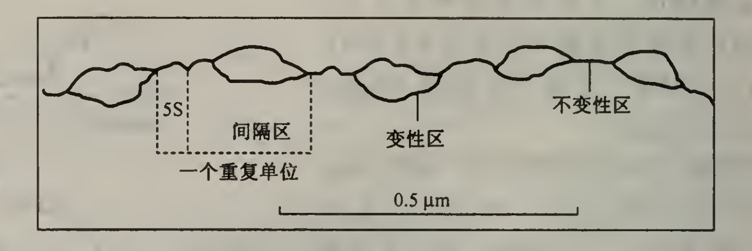


图 10-9　非洲爪蛙 5S DNA 部分变性图

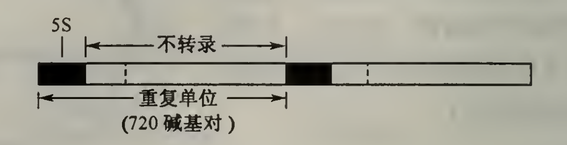


图 10-10　爪蛙 5S RNA 基因

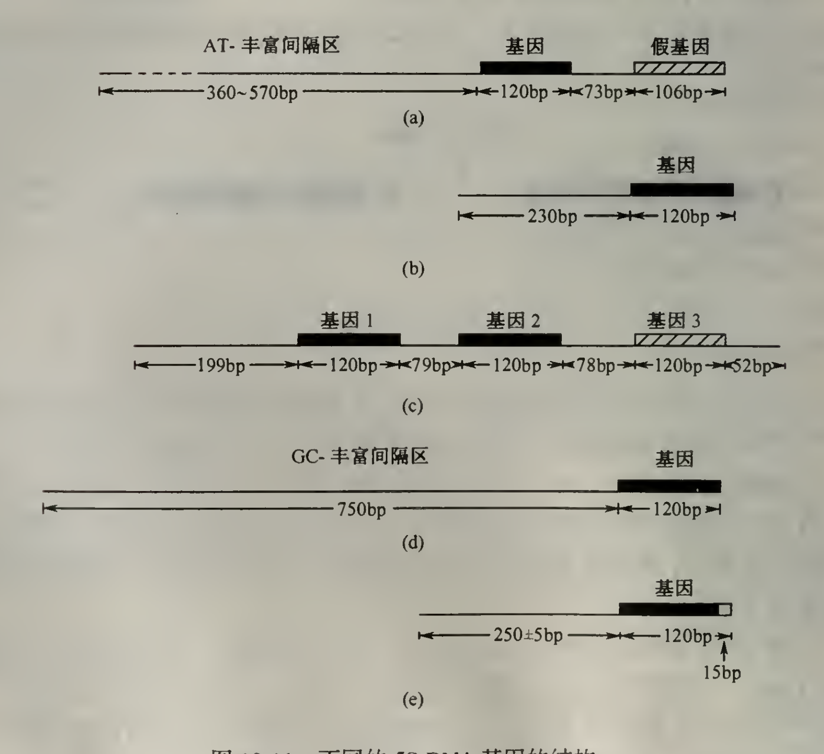


图 10-11　不同的 5S RNA 基因的结构

(a) 爪蛙大卵母细胞；(b) 爪蛙微卵母细胞；

(c) 爪蛙卵母细胞；(d) 爪蛙组织细胞；(e) 黑腹果蝇。

二、tRNA 基因

tRNA 虽小但也是细胞核酸重要的组成成分。tRNA 是一组具有广泛作用的 RNA，除了作为合成多肽链过程中氨基酸的运载工具外，还起其他方面的重要作用。例如大肠杆菌色氨酸操纵子中的弱化依赖于携带有色氨酸的 $tRNA^{Trp}$的水平，而葡萄球菌的细胞壁的合成与新的 $tRNA^{Gly}$种类相关联，在逆转录病毒基因组的 3′末端存在连有 tRNA 的结构，某种特定的 tRNA 可以具有反向转录的引物作用。这表明在高等生物中 tRNA 的功能也有类似的差异，出于不同的需要，可以估计在正常发育和癌形成期间大量不同类型的 tRNA 发生改变。因此对 tRNA 基因的研究具有广泛的意义。

1．tRNA 基因的数目

对真核生物 tRNA 基因的研究也是从非洲爪蛙开始的。用分子杂交的方法研究 tRNA 基因的结果说明爪蛙 tRNA 基因有以下特点：①tRNA 基因重复多次且各自成丛。②基因拷贝很多且基因之间有间隔区。现在已对许多种生物的不同 tRNA 基因进行研究，发现不同生物中，其细胞中 tRNA 基因总数各不相同，从几百到几千；某一种 tRNA 的基因数在不同生物中也不相同，从几个到几百个。

2．tRNA 基因的排列

在真核生物中，已知 tRNA 基因拷贝数很多，那么这些 tRNA 基因在基因组上是如何组织的便成了一个有趣的问题。它们是分散存在还是成丛存在？是每种 tRNA 基因各自成丛还是和其他 tRNA 基因一起？是否存在一个按功能的逻辑进行排列？很多实验表明真核生物中的 tRNA 基因的组织最显著的特点是多样性。

三、为蛋白质编码的基因

真核生物中为蛋白质编码的基因，研究得最快的是海胆的组蛋白基因。所有动、植物细胞染色体中都含有一类叫组蛋白的碱性蛋白，组蛋白与 DNA 形成染色体的核小体（nucleosome）。海胆染色体上的组蛋白有 5 种：H1、H2A、H2B、H3 和 H4，由 5 个基因编码，这五个基因组成一个重复单位。基因之间有间隔区，一个重复单位全长约 6500bp。在染色体上首尾相接地重复数百次。

海胆卵母细胞受精后几小时分裂 9～10 次，组蛋白也增加 2^9～2^{10}倍，所以此时的 mRNA 很容易提取。提取到其 mRNA 后，可以与其 DNA 进行分子杂交，从而提取组蛋白 DNA。用限制性内切核酸酶 *Hin*d III 处理，切口在 H1 和 H4 基因之间，因此一个 *Hin*d III 片段正好是一个重复单位。图 10-12 是一个重复单位

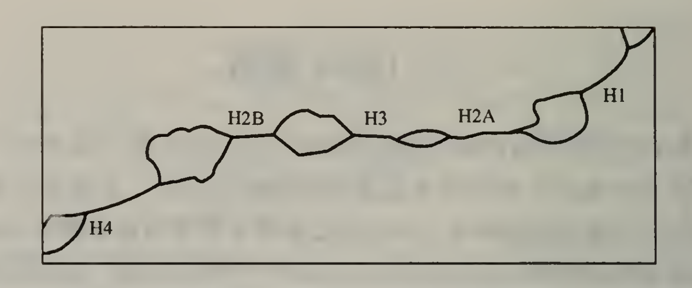


图 10-12 海胆组蛋白 DNA 一个重复单位的部分变性电镜照片

的部分变性图。

在人类细胞中，也有 5 个组蛋白基因组成一个重复单位，基因位于第七对染色体短臂上，接近于着丝点。重复数十次，人类体外培养的 Hela 细胞也是 5 个组蛋白基因组成一个重复单位。组蛋白基因的数目及排列变化很大，甚至是关系很近的生物之间都有显著的差别。表 10-1 是几种生物的 DNA 含量复制速率及组蛋白基因拷贝数的比较。从表中可以看出组蛋白基因的数目与基因组的大小，特别是与其复制速度有关。

表 10-1 组蛋白基因数目和 DNA 最大速率合成比较

生　物	DNA /(pg/细胞单倍体)	已知最短复制时间 /h	DNA 合成的最大速率 /(pg/h)	组蛋白基因拷贝数 (单倍数)
人	3	14	0.21	10～40
鼠	3	12	0.25	10～40
鸡	1.26	0.35	3.6	10
爪　蛙	4.2	0.25	16.8	20～50
海　胆	0.77	0.75	1.0	300～600
果　蝇	0.17	0.17	1.0	100
酵　母	0.025	1.0	0.025	2

以上我们介绍的都是重复多次的基因，从方法上讲，只有重复次数多的基因，才容易从细胞中直接得到，而要得到单拷贝基因则很不容易。一般是采用先获得基因的转录产物 mRNA，再把 mRNA 反转录成 cDNA。但由于基因工程技术的发展，可以将任何一目的基因在细菌中扩增，得到大量的克隆化基因，对真核单拷贝基因的研究就方便多了。

四、内部间隔区和隔裂基因

真核基因的结构，在 1977 年有重大发现，即发现了“内部间隔区”。所谓内

部间隔区是：某些基因的转录产物 mRNA 分子是一个连续不断的核苷酸共价结合的分子，但它却是从 DNA 上不相连续的片段上转录下来的。

1977 年 7 月，美国冷泉港定量生物学论文报告会上几个实验室同时报告了这一现象。当时只限于两种真核动物病毒，即猴病毒 SV40 和腺病毒，但会后很快又陆续报道真核生物本身的各种基因也具有内部间隔区，如兔和小鼠血红蛋白β链基因、鸡卵白蛋白基因、小鼠免疫球蛋白基因、酵母 tRNA 基因、果蝇 rRNA 基因以及鸡劳氏肉瘤病毒和鼠白血病病毒基因等。在为蛋白质编码的基因中，内部间隔区甚至可以位于编码部分内部。

以 SV40 病毒和鸡卵白蛋白基因为例，介绍其转录加工情况。

1. 猴病毒 SV40

SV40 是环状双链 DNA 病毒，DNA 长 5224bp，其物理图谱以 *Eco*R Ⅰ的切点为零，将全长分为 100 等份，以百分数表示，或全环为 1，等分为 100 等份后以小数表示。SV40 的环状 DNA 上，限制酶 *Eco*R Ⅰ有一个切点为零，*Hpa* Ⅰ和 *Bgl* Ⅰ共有三个切点 0.17、0.67 和 0.735（图 10-13）。

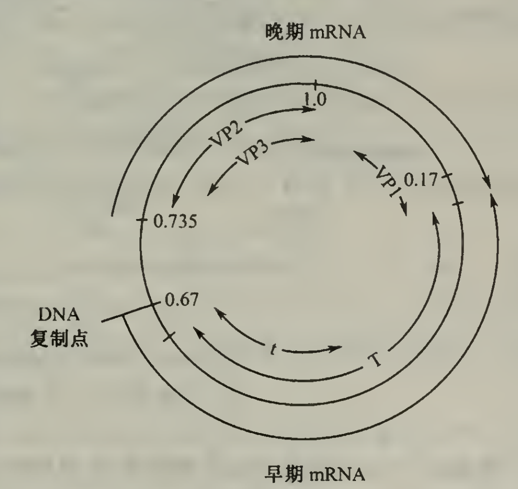


图 10-13　SV40 的物理图谱

（采自 J.D.Watson，1987）

SV40 感染宿主后，在其 DNA 复制之前就转录出 mRNA，称为早期转录。早期转录经翻译产生 T 和 t 抗原，图中 SV40 基因组环外带箭头的线代表 mRNA，环内带箭头的线表示这几种蛋白质的基因在基因组上的位置，病毒 DNA 开始复制后除早期转录外，在相反方向增加了晚期转录，晚期转录经翻译后产生三种蛋白质 VP1、VP2 和 VP3。

SV40 DNA 的转录是一个包括按时间调控（早期→晚期）和广泛的 mRNA 加工的复杂过程，这个过程中选择性地除去起始和终止密码，早期和晚期转录是在DNA 的不同链以不同的方向进行，转录的启动区靠近 DNA 复制原点（图 10-13）。图 10-14（a）表示单个早期 mRNA 分子编码 T 和 t 两种抗原，在物理图上 0.66 处开始转录，一个 AUG 密码子是两种抗原蛋白的起始密码，这个早期mRNA以两种不同的方式被加工产生单顺反子 mRNA 分子，一种产生 T 抗原，另一种产生 t 抗原。加工的模式：①切除内含子 1（intron 1）产生 mRNA

1，此 mRNA 从 AUG 起始到 Stop 1 终止翻译产生 t 抗原；②切除内含子 2（intron 2），intron 2 和 intron 1 有相同的 3′末端，切除后产生的 mRNA 从同样的 AUG 起译到 Stop 2 终止产生 T 抗原。T 抗原比 t 抗原大得多，这两种抗原的 N 端片段是相同的。

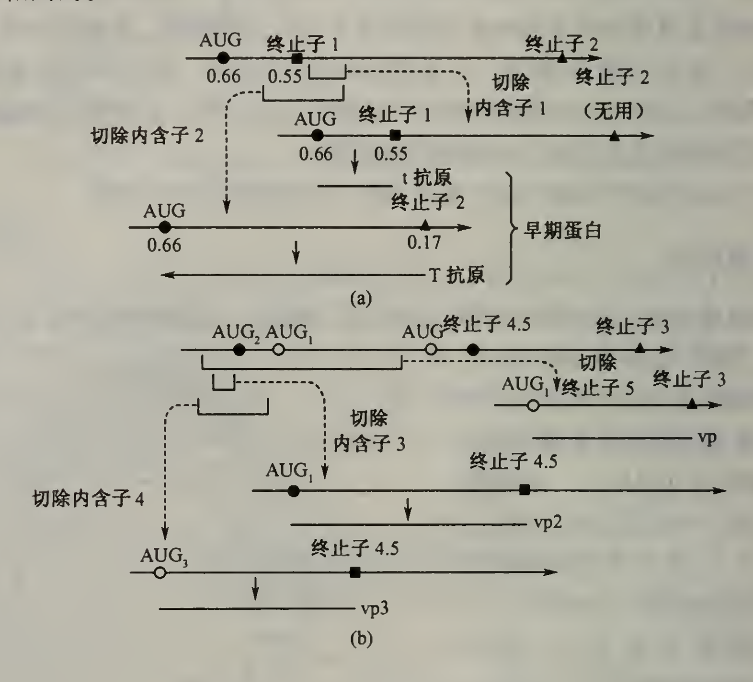


图 10-14　SV40 mRNA 的加工

（a）早期 mRNA；（b）晚期 mRNA。

图 10-14（b）表明产生三种病毒蛋白质晚期 mRNA 需三种模式的加工，没有加工的 RNA 不产生蛋白质，vp3 和 vp2 以同样的阅读框被翻译，而且二者有相同的 C 端，这样实际上 vp3 是 vp2 的一个片段；它是加工酶除去翻译 vp2 的起始密码后的 mRNA 的翻译产物。vpl 的阅读框与 vp2 和 vp3 的不同。它的阅读框是通过加工切除包括 vp2 和 vp3 的起始密码在内的很大一段 RNA 以后建立的，仅仅保留起始信号建立 vp1 的阅读框。

值得注意的是，SV40 的基因产物不是从大的多肽切割而来。

多瘤病毒（polyoma virus）DNA 的复制和 SV40 相同，其转录与 SV40 也基本相同。

2. 鸡卵蛋白基因

前述真核基因均为多拷贝基因。多拷贝基因很难研究其前后的 DNA 顺序，因此无从研究其基因的调节控制。卵蛋白基因是单拷贝基因，在鸡的各种细胞中

都存在卵蛋白基因，但是只有当母鸡下蛋时输卵管高度分化的管状腺细胞中的卵蛋白基因才表达。此基因受雌激素的控制，没有雌激素此基因不能转录 RNA，为了从分子水平上了解这一调控过程，有必要比较管状细胞和其他不表达的细胞中的卵蛋白基因的结构。法国巴斯德大学真核分子遗传学实验室的 Pierre Chambon等从上述目的出发，利用当时已经逐渐完善了的 DNA 重组技术，开始克隆鸡卵蛋白基因。

首先，他们纯化卵清蛋白 mRNA，这种 mRNA 在产卵母鸡的输卵管细胞的 RNA 中占 50%，因此易于纯化。卵蛋白 mRNA 有 1872 个核苷酸，其中 1158 个核苷酸编码卵蛋白的 386 个氨基酸。

第二步是以反转录酶转录成一条 cDNA 链，然后以 DNA 聚合酶复制成双链 DNA。

第三步是以质粒为载体，转化大肠杆菌获得 cDNA 克隆。

第四步是以限制性内切核酸酶分析 cDNA 建立物理图谱，他们注意到卵蛋白 mRNA 的 cDNA 中没有 *Eco*R I 和 *Hin*d III 两种限制酶的识别位点。

第五步是用 *Eco*R I 和 *Hin*d III 切割输卵管细胞和红细胞中分离出来的 DNA。在琼脂糖凝胶中电泳，使 DNA 片段按大小分布在凝胶板上。

第六步是用 Southern-blot 技术先使凝胶中的 DNA 片段变性，然后把它吸印到硝酸纤维滤纸上固定。

第七步是将 cDNA 用 nick translation 法制备高放射活性的 cDNA 探针，将标记的 cDNA 变性，然后与硝酸纤维滤纸上的 DNA 进行分子杂交，洗去未杂交的探针，放射自显影后即可确定含卵蛋白基因的 DNA 片段在硝酸纤维滤纸上的位置。

因为卵蛋白 cDNA 不被 *Eco*R I 和 *Hin*d III 切开，用这两种酶中任何一种切割的染色体 DNA 片段进行吸印分析，都只应该产生一条放射性带。如果染色体 DNA 上的卵蛋白基因与 cDNA 的碱基顺序是完全对应的话，卵蛋白基因中应当没有 *Eco*R I 和 *Hin*d III 的识别位点。因此，杂交只能产生一条放射性带，但是出乎意料，不论用 *Eco*R I 还是 *Hin*d III 切割的染色体 DNA 都出现数条放射性带，不论来自输卵管还是红细胞的染色体 DNA 结果都一样，这些结果反映了鸡卵蛋白基因结构中未预料到的异常现象，实验过程如图 10-15。

图 10-15 中鸡卵蛋白 mRNA 的 cDNA 与其基因组的 *Eco*R I 片段杂交产生四条带 a、b、c、d。

为了解释上述现象，必须弄清楚卵蛋白 mRNA 和染色体上卵蛋白基因 DNA 的差别。为此进一步制作 cDNA 和染色体上卵蛋白基因的限制图谱，结果证实了染色体有 cDNA 所不具有的 *Eco*R I 和 *Hin*d III 识别位点，*Eco*R I 不切割 cDNA 而染色体 DNA 被它切成 a、b、c 三段。b 段含有编码 mRNA 的起始端 500 个核苷酸顺序，而 a 段含有编码 mRNA 最后部分的顺序，从 cDNA 和染色体 DNA 的

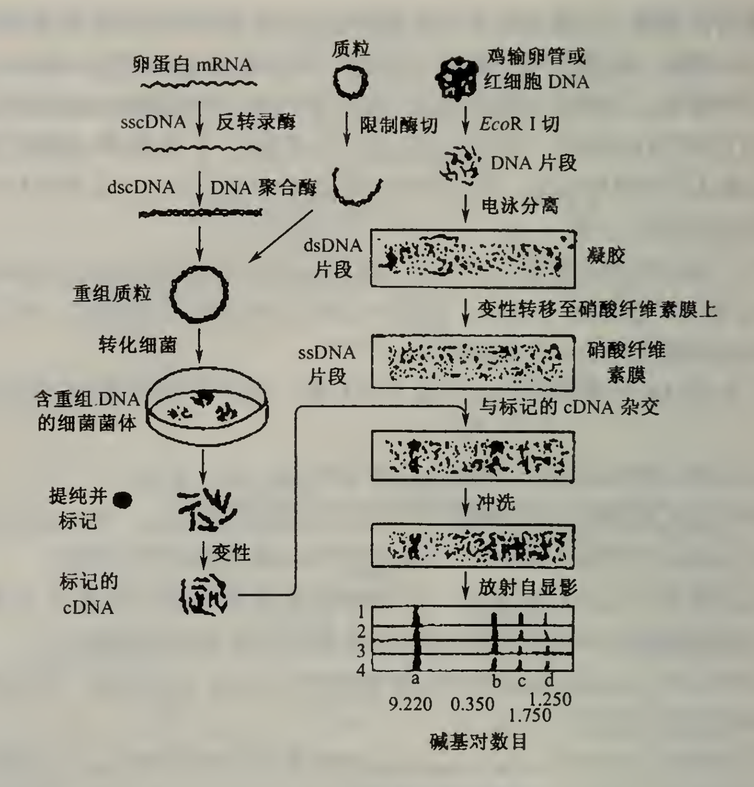


图 10-15　鸡卵蛋白基因的研究

限制图的比较，可以明显看出为 mRNA 编码的 DNA 是不连续的，这种 mRNA 最后将转译为蛋白质，也就是说为蛋白质编码的基因内是间隔的。

吸印和杂交所作的限制图谱分辨率太低，为了更详细地分析隔裂基因的结构，他们用卵蛋白 cDNA 为探针，从鸡染色体的基因文库中分离出卵蛋白基因重组分子克隆，变性后与卵蛋白 mRNA 杂交。所形成的杂种分子的电子显微镜照片（图 10-16）进一步显示出隔裂基因的结构。

从图 10-16 中可以看出鸡卵蛋白基因被 7 个内含子（A～G）隔裂为 8 个外显子（1～7），以 Maxam 和 Gilbert 方法测定 cDNA 的全部序列与已知的蛋白的氨基酸顺序完全相对应。图 10-17 中所示卵蛋白基因结构，总长约 7700bp，外显子总长为 1872bp，为基因总长的 1/4，为蛋白质编码的仅 1158bp，为基因总长的 1/7。

由于在输卵管细胞核中从未找到过卵白蛋白 mRNA 的前体 RNA (pre-mRNA)，因此有人倾向于认为卵蛋白 mRNA 是几个 RNA 分子连接而成的，那么“内部间隔区”这个名字就不适用了。因此用隔裂基因这个术语来表示这类情况。

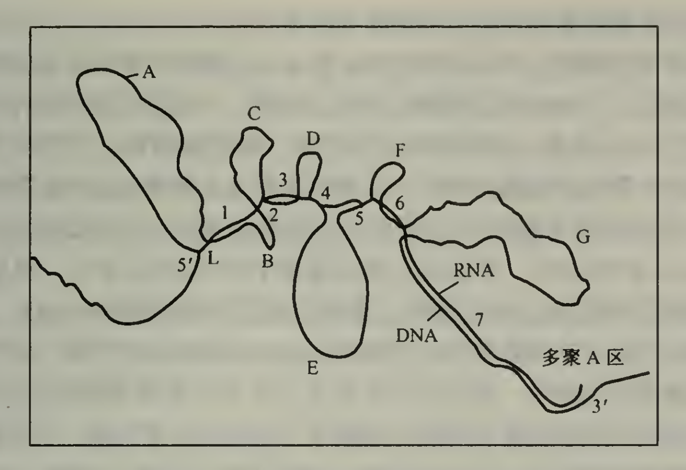


图 10-16 鸡卵蛋白基因的 intron、exon

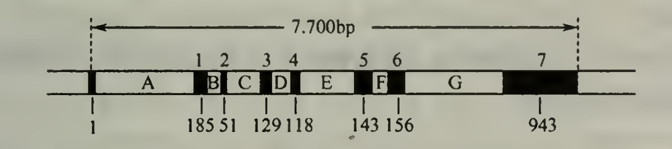


图 10-17 鸡卵蛋白基因示意图

内部间隔区和隔裂基因的发现出人意料，Pierre Chambon 等最初的设想是研究卵蛋白基因转录起始点上游发动区，以便进一步研究雌激素及其受体蛋白对这个基因的调控机制，可研究结果出乎预料，它为真核基因的研究提出了许多新问题。

第三节 真核基因表达的调控

一、真核基因表达控制的特点和复杂性

真核生物在进化上比原核生物高级，具有更加复杂的细胞结构，庞大的基因组和复杂的染色体结构。在多细胞尤其是高等动物中，随着发育和分化，出现不同的组织器官、不同类型的细胞，尽管这些不同类型的细胞的染色体具有相同的DNA，但它们合成不同的蛋白质。对不同序列的 mRNA 研究表明，较高等的真核细胞中合成约 10 000～20 000 种蛋白质，其中含量丰富的蛋白质大部分在不同类型的细胞中相同，而至少有好几百种含量低的蛋白质在不同类型的细胞中是不同的，其中大部分是调控蛋白或酶，这几百种不同的蛋白虽然占的比例不大，但已足以使细胞出现不同的形态和行为，行使不同的功能。和原核生物相比，真核

生物的基因表达调控更加复杂，更加精细和微妙。

真核基因表达调控，可以在从 DNA 到 RNA 到有功能的蛋白质的途径中的许多步骤上进行，①DNA 和染色体水平上的调控：基因的拷贝数的扩增或丢失和基因重排，DNA 修饰，在染色体上的位置、染色体结构（包括染色质、异染色质、核小体）都可影响基因表达；②转录水平上：转录起始的控制和延伸的弱化对 mRNA 前体的水平都会产生影响；③转录后 RNA 加工过程和运送中的调控：原核 DNA 被转录时，蛋白质合成也随着转录起始而进行，即边转录边翻译，而真核基因转录出 mRNA 前体，要经过加工才能成熟为 mRNA，包括切割、拼接、编辑、5′和 3′末端修饰等，成熟的 mRNA 再运出细胞核；④翻译水平上的调控；⑤翻译后的控制：翻译后产生的蛋白质常常需要修饰和加工如磷酸化、糖基化、切除信号肽及构象形成等，才能成为有活性的蛋白质，可以利用这个过程有选择地激活或灭活某些蛋白质；⑥mRNA 降解的控制：控制 mRNA 的寿命就能控制一定数目的 mRNA 分子产生蛋白质的数量，这种控制是由 mRNA 3′端的序列决定的。

上述归纳的几个方面的控制，最重要的是转录起始的调控和翻译水平上的调控，我们下面将重点介绍。

二、真核基因转录起始的控制机制

以前所述的在原核细胞内转录控制中，乳糖操纵子的两种信号调节机制是简单而完美的，当起负控制作用的抑制蛋白从 DNA 的结合区脱离后，便让出空间位置，使 RNA 聚合酶得以和 DNA 结合而起始转录，但在真核细胞中转录起始情况就完全不相同，真核基因转录激活和起始需要多种元件，包括特定的 DNA 序列、蛋白质激活因子、通用转录因子和 RNA 聚合酶，也就是说转录激活和起始涉及 DNA 和多种蛋白质之间相互作用的大分子体系，要弄清真核细胞如何精心创造出如此复杂的基因开关，我们必须对下面叙述的问题逐一了解。

1. 真核基因控制区包括启动子和调控的 DNA 序列

真核基因控制区指的是起始转录所需的 DNA 序列，加上调控起始发生的速率所需的 DNA 序列。因此，控制区包括启动子（promoter）和调控序列。

启动子是通用转录因子和 RNA 聚合酶装配成转录起始复合物的一段 DNA 序列，对它下游的一个转录单位的转录起正控制作用，这段序列对于 RNA 聚合酶 II 的启动子来说，位于转录起始点附近至上游的 TATA 盒之间，RNA 聚合酶 III 转录的基因有两个启动子：一个在上游，另一个在转录序列的中间。几乎所有的 RNA 多聚酶 II 利用的启动子都有 TATA 序列，TATA 盒一般位于相对转录起始核苷酸上游 −25 区，是启动子的标志。

决定转录起始速率的调控序列，是一些调控蛋白，包括转录激活蛋白和抑制蛋白的结合区域。一般一种激活转录的调控蛋白（反式激活因子）识别和结合的DNA长度在20个核苷酸之内，虽然一般在启动子上游200个核苷酸之内存在这些序列，但在较高等的真核生物中，调控序列点布在大到50kb的范围之内并非罕见，增强子上含有激活转录的调控蛋白结合序列，能增强上下游远距离的基因转录。在RNA多聚酶II转录的基因控制区，常见的短序列有TATA盒、CAAT盒和GC盒。

TATA盒位于转录起始点上游-25左右，是通用转录因子TFIID识别和结合的序列，TATA盒发生变异对转录水平影响不大，但影响转录的起始点，一致性序列为TATAAA。

CAAT盒：一般位于-75左右，一致性序列为GGGCCAATCT，是转录激活因子NF1和CTF成员等多种蛋白识别和结合序列，对基因转录有较强的激活作用，对两个方向都可激活，而且作用距离不定。

GC盒：GGGCGG，是转录激活因子SP1识别与结合序列，它和CAAT盒都是不对称的序列，但都能对两个方向行使激活功能，GC盒的位置可以处于TATA盒与CAAT盒之间，也可位于CAAT盒上游，位置因不同基因而异。

上面说到的几种短的DNA调控序列只是常见的DNA调控序列的一部分，事实上，不同基因控制序列千差万别。这些短的DNA调控序列是基因开关的基本成分。

2. 真核基因转录调控蛋白识别和结合特异的DNA序列

真核细胞中的转录调控蛋白有好几千种，种类繁多，但每种在一个细胞中的含量都非常少，种类依不同细胞类型而异，不同基因受不同的调控蛋白控制，这样能允许细胞启动或关闭单个特异基因的转录。

转录调控蛋白，一般能识别和结合DNA上的一段短的序列。这段短的DNA序列可以用足迹试验（foot printing）和凝胶迁移率改变试验（gel-mobility shift assay）来测定。通常说的"顺式激活因子"（cis-activative factor）指的就是激活转录作用的调控蛋白识别和结合的一段DNA序列，它只对同一DNA上的基因具激活作用，也有人把它称为同域作用元件。而"反式激活因子"（trans-activative factor）就是能激活基因转录的调控蛋白，它们对处在不同DNA分子上的基因都能起作用，也有人称为异域作用因子。事实上，"反式"二字并不合乎英文的原意，原意应为"转移性激活因子"。表10-2中列举了一些基因转录调控蛋白及其识别的DNA序列。也有的调控蛋白本身并不直接识别和结合于DNA的特异序列，而是通过与结合在DNA上的其他调控蛋白结合，通过蛋白质之间的互相作用而加强或抑制转录，例如一些病毒的最早期基因产物（IE蛋白）可以激活许多种基因的转录，这些基因的控制区DNA序列并不相同。

表 10-2　一些调控蛋白及识别和结合的 DNA 序列

调控蛋白		识别结合序列	分子质量或功能	
SP1		GGGCGG(GC box)	95～110kDa;	
OCT1,2/(OTF1,2)		ATGCAAAT ATTTGCAT	OCT1 90～100kDa;OCT2	
GATA-1		TGATAG		
CTF/NF1		GGCCAATCT(CAAT box)	抗原上与 CP-1 无关	
NF-κB	65.50kDa	GGGACTTTCC		
ATF		GTGACGT		
TBP		TATAAAA(TATA box)		
AP1	39kDa (p^{39})	TGACTCA		
AP2	50～52kDa	CCCCAGGC		
TF II D	32kDa	TATA	32kDa	
E2F		TTTCGCGC		
CRE	35～43kDa	TGACGTCA(cAMP 应答因子)	cAMP 诱导	L Zip
SRF	67kDa	GATGTCCATA		
C/EBP	42kDa	TGTGGAAAG(增强子结合蛋白)		L Zip
C-myc		CCGGGTTTTCCCC		
C-Jun		TGAC/GTCA		L Zip
GAL4		CGGAGGACTGTCCTCCG		
CP-1		CCAAT		

3. 基因转录激活蛋白的功能区域和结构花式

许多基因转录激活蛋白可分为明显的两个功能区域，一个是识别和结合 DNA 的特异序列，另一个是激活转录的功能区域，后者无基因特异性，如果把 A 激活蛋白的 DNA 结合部分与 B 激活蛋白的激活功能部分融合在一起，可以构成一种有功能的基因激活蛋白，这种融合蛋白只能激活 A 蛋白所能激活的基因。激活的功能区域负责与其他的蛋白质互相作用，如形成二聚体，帮助通用转录因子在启动子上装配成转录起始复合物等。这一功能区域在结构特点上常在表面富含带负电荷的酸性氨基酸，被称为“酸性的激活区域”。

在生物学上分子识别一般依靠两种分子表面之间能完全配套。基因调控蛋白识别双链 DNA 上的特异序列是因为蛋白质表面与 DNA 双螺旋的特殊的表面通过氢键、离子键和疏水的互相作用形成广泛的互补，例如门冬氨酸可和 A 结合，精氨酸和组氨酸可以和 G 结合，识别 DNA 序列的蛋白质结构主要是 α 螺旋，但

也有β折叠。对于被识别的双链 DNA 来说，并不需要解开螺旋和形成单链，每个碱基对的边缘是暴露在双螺旋的表面的，能为蛋白质的识别呈送不同的氢键供体、氢键受体和疏水的模式。

基因调控蛋白含有能阅读 DNA 序列的结构花式（motif)。几种主要的结构花式及其与 DNA 的作用方式见第二章。

4. 真核 RNA 聚合酶需要的通用转录因子

原核 RNA 聚合酶可以单独识别和结合在原核基因的启动子上而起始转录，变换不同的σ亚基就能转录不同套的基因。真核 RNA 聚合酶不同，它在体外不能起始转录，于是导致发现另一些称为“通用转录因子”的蛋白质。通用转录因子并非是真核 RNA 聚合酶在被提纯过程中丢失的亚基，它们是一些独立的蛋白质。每种 RNA 聚合酶转录任何一个基因使用的是同一套通用转录因子，每种转录因子含多种蛋白质亚基。与调控蛋白相比，转录因子种类少，但含量丰富。真核RNA 聚合酶 II 使用的一套转录因子包括 TFIID、TFIIA、TFIIB、TFIIF、TFIIE、TFIIH 和 TFIIJ。

转录因子和 RNA 聚合酶 II 必须在启动子上装配成复合物才能起始转录，装配过程如前所述。

5. 增强子通过多蛋白复合物而增强转录

增强子的长度在 50～200bp 范围之内，内含一系列转录因子的结合位点。对人类β-干扰素基因增强子（大约 70bp）分析，为与增强子相互作用的转录因子的作用模式提供了一个很好的例子。通过接头扫描突变（linker scanning mutation）分析，已鉴定出四个控制元件和与每个控制元件识别结合的蛋白。这些蛋白在转录实验中激活来自β-干扰素增强子的转录。图 10-18 转录实验显示这些转录因子对β-干扰素增强子的结合是同时发生的，在一个与染色质联系在一起的丰富的小蛋白质 HMG1 存在情况下，转录因子的结合具高度协同性，与 *E.coli* 中 CAP 蛋白和 RNA 聚合酶在启动子上的结合（图 10-19）类似，这种协同结合在增强子 DNA 上产生一个蛋白质复合物，命名为增强体（enhancesome）（图 10-20)。

HMG1 结合于 DNA 小沟，使 DNA 分子急剧弯曲，这种弯曲使转录因子可适当地相互作用。因为这些结合蛋白结合在相邻的位点，使蛋白质的局部浓度增高从而它们之间的相互作用增强。在研究 T 细胞受体基因α的增强子中，同样发现一种类似于 HMG1 的功能的蛋白质也是与 DNA 小沟结合并诱导 DNA 弯曲。这类蛋白被命名为 architectural protein，因为它对构建这些多蛋白复合物是必需的。

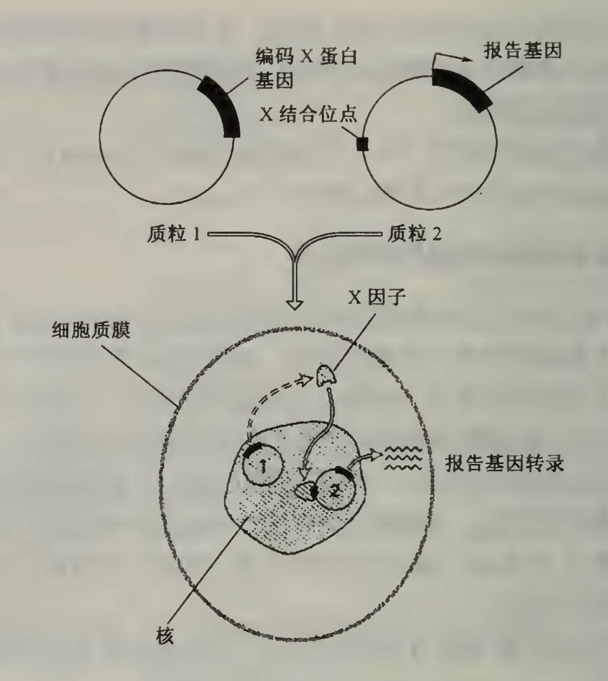


图 10-18　转录因子活性的体内分析

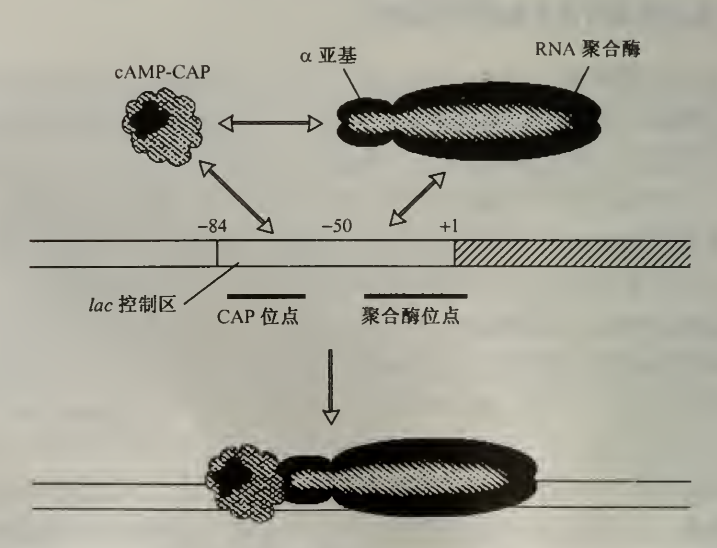


图 10-19　*E.coli* RNA 聚合酶和 cAMP-CAP 对 *lac* 启动子的协同结合

6. HIV 基因组的转录通过抗终止被调控

通过对 RNA 聚合酶 II 进行 HIV 基因组转录的研究提供了真核中转录终止调控的例子。HIV 基因的有效表达需要 *tat* 基因编码的病毒小蛋白。*tat*$^{-}$突变株

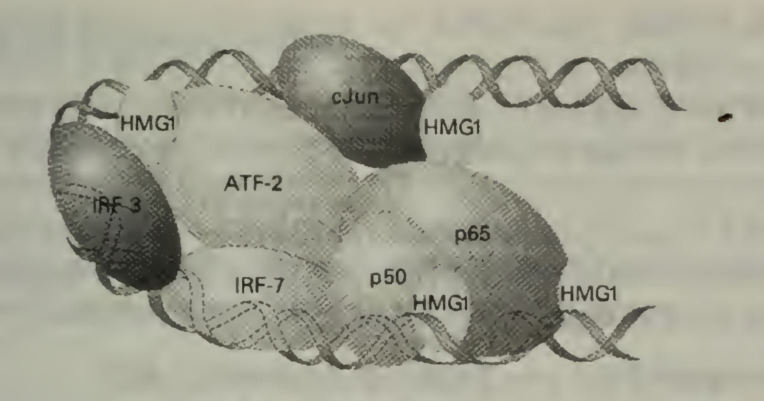


图 10-20　在 β-干扰素增强子上形成的增强子体（enhancesome）的模型

感染的细胞只产生短的 HIV 转录子，它只能与 HIV DNA 上靠近启动子的限制片段杂交，而不能与远离启动子的限制片段杂交。而在野生型 HIV 感染的细胞中，合成长的病毒转录子，它能与 HIV 的全长 DNA 中任一限制片段杂交。由此看来，Tat 蛋白的功能是作为抗终止因子，使 RNA 聚合酶 II 可通读转录封闭区，类似于 λ 噬菌体的 N 蛋白，Tat 蛋白对于 HIV 复制也是必需的。进一步了解这种基因控制的机制，能为设计艾滋病的有效治疗药物提供可能。

Tat 是顺序专一性 RNA 结合蛋白，它结合在 HIV 转录子 5′末端的一个称为 TAR 的顺序上，该顺序也类似于 λ 转录子的 nut 顺序。TAR 含有两个结合位点，一个与 Tat 蛋白相互作用，另一个与称为 Cyclin T 的细胞蛋白相互作用。nut 和 RNA 结合域也与 N 蛋白类似，含有精氨酸丰富区。

图 10-21，HIV Tat 蛋白与 Cyclin T 对 TAR 的结合类似于 λ N 蛋白对 *E. coli* Nus B 和 S10 与 nut RNA 的相互作用（图 9-25）。Cyclin T 与蛋白激酶 Cdk9 相互

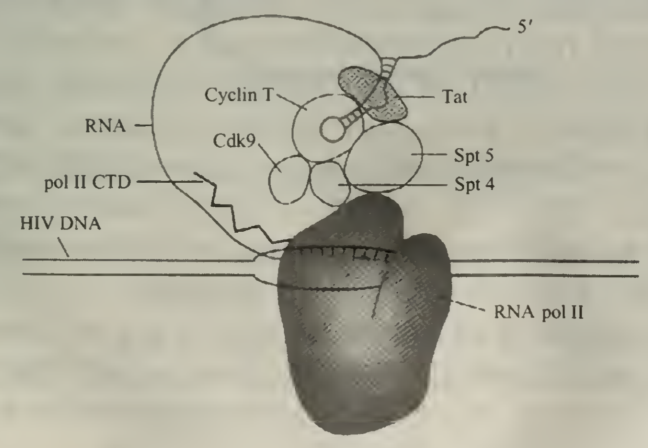


图 10-21　HIV Tat 蛋白和几种真核细胞蛋白组成的抗终止复合物

作用，激活该激酶，Cdk9 的底物是 RNA 聚合酶 II 的 CTD。在体外转录研究中，加入 Cdk9 的抑制物，除了 Cdk9 高度磷酸化外，RNA 聚合酶 II 在 HIV 启动子起始转录大约 50 碱基后终止。Cyclin T 和 Tat 与 HIV 转录子的 5′端的 TAR 顺序共同结合，位于能使 CTD 磷酸化的 Cdk9 处，抗终止而允许 RNA 聚合酶进行链延伸。

很明显，Cyclin T-cdk9 复合物中的一个蛋白质——Spt 5 具有与 Nus G 蛋白的氨基酸顺序同源区域，Nus G 蛋白是 λ N 蛋白抗终止中的一个相关蛋白。这些同源性暗示从细菌到动物，在调节 RNA 链延伸中具有类似的机制。

7. 某些快速诱导基因中 RNA 聚合酶 II 的近启动子处发生暂停

果蝇的热休克基因（heat-shock gene，如 hsp70）代表了真核中 RNA 链延伸调控的另一种机制。

在果蝇热休克基因的转录中，RNA 聚合酶 II 起始转录大约 25 个碱基后暂停，而不是终止（如 HIV 基因组在缺失 Tat 蛋白时）。被暂停的 RNA 聚合酶 II 仍保持与新生 RNA 和模板 DNA 的联系。直到热休克转录因子 HSTF（heat shock transcription factor）的激活这一条件的发生，紧接着被激活的 HSIF 结合到热休克基因近启动子区的专一位点，促使被暂停的 RNA 聚合酶 II 继续 RNA 链的延伸，并且快速启动新加入的 RNA 聚合酶 II 的重新起始。

热休克基因是由细胞内蛋白质变性条件（诸如温度升高“heat-shock”）诱导的。一些热休克基因是编码较抗变性条件和起防止其他蛋白变性作用的蛋白；另一些热休克基因是编码帮助变性蛋白质重新折叠的伴侣。

这种包含调控这些基因表达的转录控制机制允许一个快速的应答，因为不需要时间去装配转录起始复合物。

转录起始复合物在 RNA polII 启动子上装配的过程，每一步都可能受调控蛋白的影响，尤其是 TFIID 与 TATA 的结合和 TFIIB 与 TFIID 的结合，即使在体外用提纯的 DNA（无组蛋白、核小体、异染色质和染色质结构）作为转录模板，加入提纯的通用转录因子和 RNA polII，转录效率仍是很低的。如果 DNA 的启动子上游有激活转录的调控蛋白结合序列，加入激活转录的调控蛋白质后可以使转录水平提高几十倍至几百倍。这就说明转录起始复合物装配速度和频率受调控蛋白的控制，激活转录的调控蛋白帮助和加速转录因子装配成转录起始复合物。当 RNA 聚合酶从复合物释放后转录因子复合物分散解体。

真核 RNA 聚合酶 I 和 III 也需要自己独特的一套转录因子，RNA polI 需要 UBFI 和 SL1，SL1 含 TBP 和另外三种蛋白。RNA polIII 需要 TFIIIA、TFIIIB、TFIIIC，TFIIIB 含 TBP，TBP 在 polI 和 pol III 启动子上的结合不依赖 TATA 序列。

8. 真核基因转录抑制蛋白的作用模式

上面我们主要介绍了激活转录的因子，基因控制包括激活和抑制两个方面。起抑制作用的调控蛋白，有的通过和特异性 DNA 序列结合后抑制基因转录，有的在细胞质中与激活转录的调控蛋白结合形成二聚体而使激活转录的调控蛋白成无活性状态。例如 I-κB 与 NF-κB 结合成二聚体使 NF-κB 灭活，只有当 NF-κB 从二聚体中释放出来后才能进入核中与 DNA 结合激活转录。与 DNA 结合后抑制转录的调控蛋白的几种作用方式如图 10-22 所示。

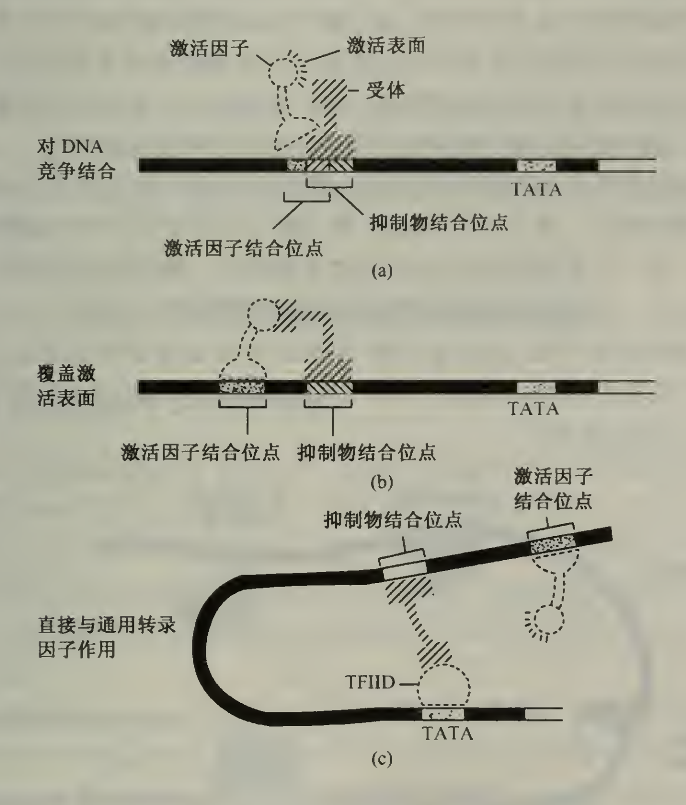


图 10-22　真核基因抑制蛋白可以三种方式起作用

1）竞争结合：激活蛋白与抑制蛋白的 DNA 结合序列有部分重叠，如抑制蛋白先结合占领了空间位置，激活蛋白就不能结合。

2）封闭或修饰激活蛋白的表面：抑制蛋白结合在激活蛋白近旁或通过 DNA 弯曲回折使抑制蛋白与激活蛋白靠拢，抑制蛋白与激活蛋白的激活转录功能区域相互作用而使后者失去激活转录活性。

3）直接与通用转录因子或 RNA 聚合酶相互作用阻止转录起始复合物的装配。

4) 结合在起始复合物装配的区域或下游，阻止 RNA 聚合酶转录，例如 HCMV IEI 蛋白是一种转录激活因子，可激活多种基因的转录，但 IEI 蛋白对自身基因的转录具有自调作用或反馈抑制作用，IEI 蛋白可与自身基因 DNA 结合，结合位置在转录起始附近。

9. 真核基因转录调控的一般模型（DNA 和多种蛋白质相互作用）

上面已经逐一叙述，真核基因转录控制包括三大要素：①DNA 调控单元：DNA 上一些短的序列，是转录因子和调控蛋白结合的区域，这些区域一般分布在转录起始区上游，如 TATA 盒、GC 盒、CAAT 盒、增强子及其他序列。②调控蛋白：如 SP1、NF1、OCT、NF-κB 等。③通用转录因子和 RNA 聚合酶。这三大要素的大分子系统相互作用包括 DNA 与蛋白质相互作用、调控蛋白之间的相互作用、调控蛋白与通用转录因子及 RNA 聚合酶的相互作用。

在大分子体系的相互作用中，DNA 调控单元起了提供相互作用的场所和组织者或导演的作用，图 10-23 是 RNA 聚合酶 II 转录的综合性的调控模型，在这一模型中，第一步是调控蛋白与 DNA 的互相作用，激活转录的调控蛋白与 DNA 特异序列结合，激活转录的蛋白帮助和促进 TFIID 进入和结合 TATA 盒，激活蛋白与其他转录因子和 RNA 聚合酶 II 相互作用加速转录起始复合物的装配。TFIIF的大亚基使 DNA 解螺旋，TFIIH 的作用使 RNA polII 从复合物中释放，转录产生 mRNA 前体。

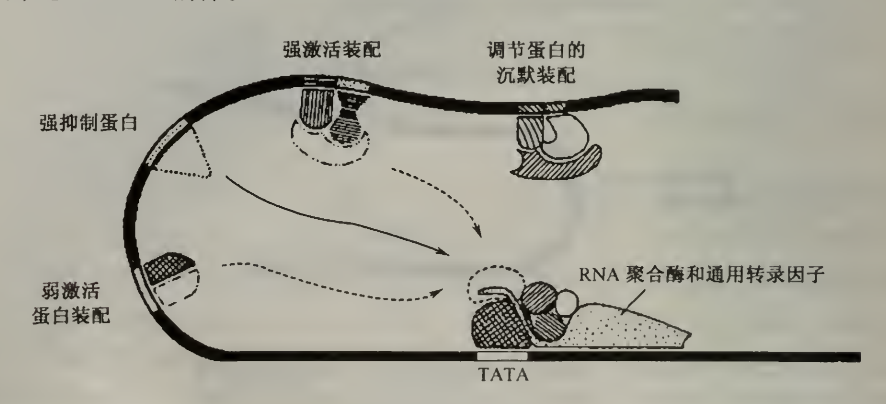


图 10-23　在启动子处各种调控因子的相互作用

模型中有多种激活转录蛋白与不同的 DNA 序列结合，每种能单独激活转录，但每种激活的强度有差别，多种蛋白（包括增强子上结合的蛋白）协同作用能进一步增强转录。另外值得注意的是，有的 DNA 序列可以被不同的激活因子识别和结合，也有的调控蛋白可识别和结合两种以上的 DNA 序列，例如 C/EBP。不同因子结合同一序列对转录激活的强度不一样，例如 OCT 序列可以被

OCT1、OCT2 识别与结合，CAAT 盒可被 NF1、C/EDP、CTF 的成员、ACF、CDP 等多种因子识别与结合。各种因子的结合亲和力不同，对转录激活的作用也不一样，当 CDP 结合在 CAAT 盒时可以抑制转录。不同的组织，不同类型细胞，不同发育时期出现不同的调控蛋白，甚至同一细胞的不同基因的相同 DNA 序列，可能结合的调控蛋白也有区别。但对多数 DNA 序列来说结合的调控蛋白是相当专一的。

从上面的模式图中我们看到，在真核基因的控制区有许多不同的调控蛋白（激活转录的和抑制转录的）结合。虽然在体外的 CAT 试验中，每种激活转录的调控蛋白结合在 DNA 上通常能单独激活转录，可以测出报告基因的表达，但在真核细胞中，启动子上的转录是开启还是关闭，转录是强还是弱，必须集中所有结合的调控蛋白提供的对转录的提示，启动子才能做出决定。增强子是远距离控制单元的代表，它能使转录水平提高几十倍甚至几百倍，通过 DNA 回折靠拢启动子，下面我们再介绍几种具体的真核基因转录调控模型，以便加深对真核基因转录调控的理解。

三、几种真核基因转录调控模型

1. 果蝇胚胎 *eve* 基因条纹 2 的调控模型

果蝇 *eve* 基因是胚胎发育早期（单个多核巨细胞时期）的重要基因，如 *eve* 基因灭活则胚胎的许多部分不能形成，胚胎就在发育早期死亡，它的表达控制区长 20kb 以上，结合 20 多种调控蛋白。这一巨大的复杂的调控区含一系列 DNA 调控单元，每单元排列多个独特的短的 DNA 序列，这些短序列为特异的调控蛋白结合区。目前对整个调控区的研究还不大深入，我们下面只介绍和讨论条纹 2 单元的调控模式。

条纹 2 控制单元长 480 个核苷酸，已知排列有 8 个蛋白质结合序列，可以以不同方式结合四种调控蛋白，其中 bicoid 和 hunch hack 两种蛋白具正控制功能，krüppel 和 giant 是两种抑制蛋白，这四种蛋白共同控制 *eve* 基因转录的开关，两种抑制蛋白中任一种结合于 DNA 上就能关闭条纹 2 控制单元控制的转录，如果果蝇的这两种抑制蛋白缺乏，受精的果蝇卵（胚胎）的组织化学染色出现的条纹 2 变得很宽。而要使控制的基因达到最大的表达，两种正控制的调控蛋白必须同时都结合在 DNA 上。

从图 10-24 中可以看出，激活蛋白和抑制蛋白的结合区排列有两个特点，一个是 8 个结合序列中有 4 个激活蛋白和抑制蛋白的 DNA 结合序列重叠；另一个是激活蛋白的结合序列被抑制蛋白的结合序列隔开。至于为什么两种抑制蛋白中任一种结合于 DNA 序列上就能关闭转录还不清楚，因为目前还不了解它们相互

作用的细节。如果把 *lacZ* 基因接在条纹 2 调控单元的下游，可以用 DNA 突变技术研究调控单元中各个短序列或每种控调蛋白对基因表达的影响。把这种人造的杂合基因转移到果蝇早期胚胎的染色体中，组织化学染色可以使条纹 2 变为蓝色（lacZ 的缘故，图 b 中箭头所示），条纹的位置不改变见图 10-25。

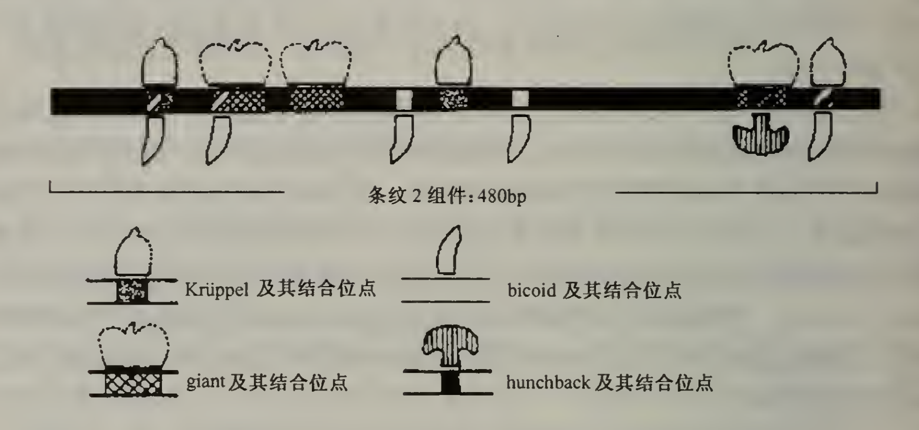


图 10-24　条纹 2 单元调控序列与调控蛋白的相互作用

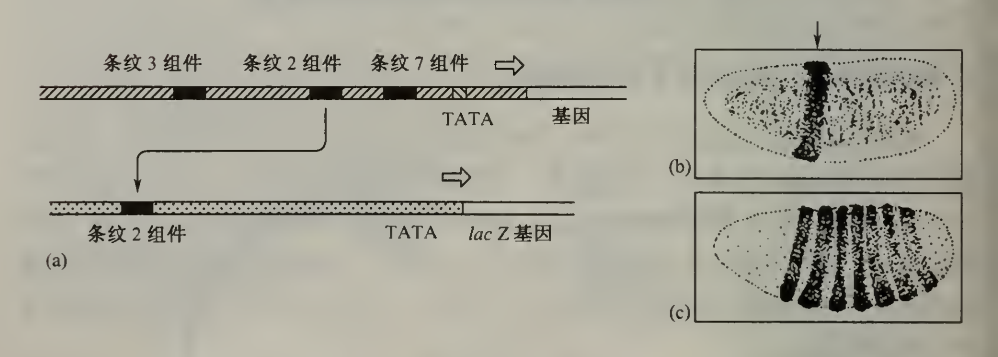


图 10-25　实验阐明的 *eve* 基因调控区的标准元件结构

2. 人 β 球蛋白基因控制模型

人 β 球蛋白基因控制是复杂的真核基因控制的典型之一，也是目前了解较为详细的真核基因控制模型。β 球蛋白基因只在红细胞发育的特定时间表达。因此，β 球蛋白基因表达具有细胞类型和发育时间的专一性。

β 球蛋白和其他球蛋白基因在染色体上排列成簇组成基因座（gene locus），连同基因座的控制区，跨越约 100kb 的 DNA 长度，基因座中每个球蛋白基因除有自己的控制序列外，整个基因座还有制控序列，称为 LCR（locus control region），基因座中各球蛋白基因排列次序和具有不同 5′LCR 的球蛋白基因在不同

发育阶段和不同器官中表达。

β 球蛋白基因及其本身的控制区如图 10-26 所示。转录区域长近 2kb，转录区的上游和下游各有一段长约 200bp 的调控序列，两个控制区对 β 球蛋白基因的最大表达都是必需的，但上游的控制区对转录的开关起关键作用。

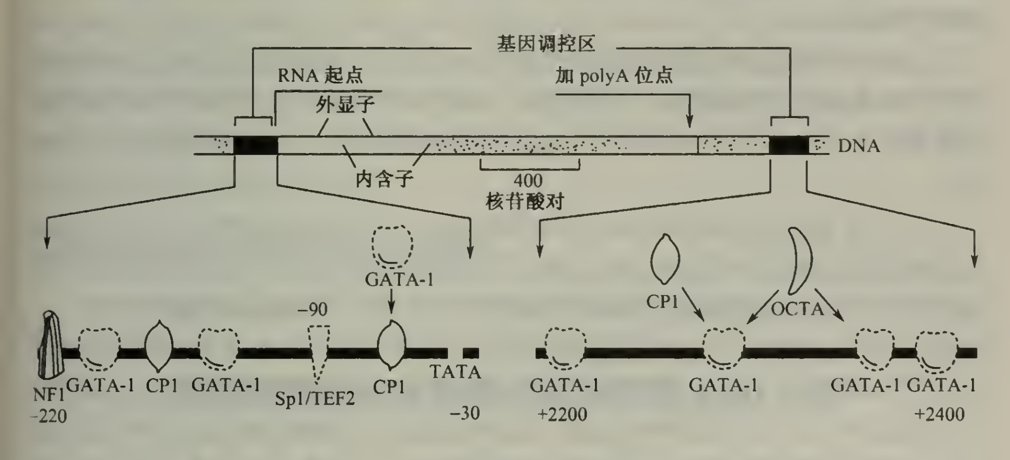


图 10-26　人 β 球蛋白基因的调控模型

下游的控制区比较简单，含四个 GATA-1 的结合区，但其中的两个结合区不是 GATA-1 专一的，也可以结合 OCTA 或 CP1（结合序列重叠，箭头所示）。上游的控制区显得较为复杂，在 TATA 盒的上游有 6 个短的序列可以结合四种不同的调控蛋白。在这些调控蛋白中，SP1、CP1、NF1 在一般的细胞类型中都存在，而 GATA-1 这种调控蛋白只在包括红细胞在内的很少几种细胞中存在。GATA-1 识别和结合的 DNA 核心序列为 TGATAG，是 β 球蛋白基因激活的重要蛋白，可见在发育分化过程中，调控蛋白的种类是在不断改变的。

如果把人的 β-球蛋白连同它的近旁调控区一起转移到转基因老鼠中，无论其插入在染色体什么位置，表达的效果都是很差的，这说明 β-球蛋白基因除受本身上下游的调控序列调控外，还受到更远的调控序列控制。研究结果表明，球蛋白基因簇上游的染色质结构控制整个基因簇的转录开关。第一个证据来源于核酸酶敏感试验。

在球蛋白基因不表达的细胞中，在分离的细胞核中的球蛋白基因对DNase I 有抗性，但在红细胞分离的核中的整个球蛋白基因簇都对 DNase I 超敏感，表明在前者中基因被紧密压缩包裹在染色质中，而后者染色质变得松弛而使 DNA 某些部分暴露出来，这暴露出来的 DNA 序列被认为是激活转录的调控蛋白结合区，当调控蛋白与这些序列结合后进一步引起染色质核小体改变而开启各个基因的转录。在 ε 球蛋白基因上游约 20kb，发现 5 个 DNaseI 超敏感区，这些敏感区就是 LCR，LCR 序列不但对 β 球蛋白转录是必需的，而且对基因簇中的其他球

蛋白基因的转录也是必需的。

LCR 的调控作用还可以从对地中海贫血病（一种球蛋白生成障碍的贫血病）的研究中看出，病人的 β 球蛋白及近旁的调控序列完整无缺，这部分也对 DNase I敏感，但 LCR 部分或全部缺失，因此使 β 球蛋白基因在病人的红细胞中保持休眠状态。

LCR 的作用在对转基因老鼠研究中进一步证实，把 β 球蛋白及其近旁的调控序列插入鼠红细胞基因组中不同位置，表达水平因插入位置不同而有些差异，但都是低水平表达。当加上 LCR 一起，无论在哪里插入，都能获得高水平的表达。

与 LCR 结合的几种调控蛋白已被鉴定出来，但问题的关键是 LCR 的 DNA 与调控蛋白的相互作用怎样改变下游广大范围的染色质结构？目前虽然有几种假设，但对于 LCR 远距离调控的机制和细节还不了解，还不能做出结论。

四、DNA 修饰和染色质结构控制基因转录

1. DNA 上特异位置甲基化可控制基因表达

我们在前面主要讨论 DNA 的调控序列、启动子、调控蛋白、对真核基因转录的控制，事实上，真核基因转录水平的控制是多方面的。

在真核细胞中，DNA 上甲基化的基因主要集中在异染色质，有 2%～7% 的 C 被甲基化。事实上大多数甲基化在 CG 二联序列，在两条链上甲基化产生 $5'{}^{m}CPG\ 3'$，这种甲基化可用限制性内切核酸酶分析检查。*Msp* I 能消化所有的
$3'\ GPC^{m}5'$
CCGG 序列，而 *Hpa* II 只能切割未甲基化的 CCGG 序列，两种酶切割比较就知道有多少 CCGG 被甲基化。甲基化图式检查发现，许多基因中大多数的甲基化位置是比较固定的，但另有一些位置是可变的。基因在不表达的组织中大多数位点是被甲基化的，但在基因表达活跃的组织中是没有甲基化的。把甲基化的和不甲基化的基因导入新的宿主细胞中，结果甲基化的基因不表达，未甲基化的基因可以表达。这表明了某些基因的特异位点 CG 甲基化可抑制转录，对这些基因转录必须经过去甲基化过程。那么，什么地方的 CG 甲基化才能抑制转录？甲基化抑制的机制是什么？

γ 球蛋白基因在转录起始点周围即 －200～＋90 之间存在多点甲基化，无论去除起始点上游三个位点的甲基化或去除起始点下游三个位点的甲基化，都不能解除转录的抑制，当除去所有位点的甲基化后，γ 球蛋白基因才能被转录，因此认为启动子的甲基化是抑制转录的，其余广泛的地方甲基化不抑制转录。在一些甲基化影响转录的基因的 5′区域常存在 CpG 丰富岛（CpG-rich island），这些地

方的 GC 含量比 DNA 的 GC 平均含量明显高得多，而 CG 二核苷酸序列出现的频率比基因组的其他地方增加 10 倍以上。

在启动子的 CpG 岛甲基化防止基因转录是由两种结合蛋白造成的，这两种蛋白是 MeCP-1 和 MeCP-2，MeCP-1 的结合需要几个 CpG 甲基化，MeCP-2 的结合只要一个 CpG 甲基化。两种 CpG 甲基化结合蛋白的任一种结合于启动子都能抑制转录。MeCP-1 和 MeCP-2 在启动子的结合阻止通用转录因子和 RNA 聚合酶在这一位置装配成转录起始复合物。由此可见，某些基因的转录激活还包括 DNA 去甲基化步骤。甲基化控制基因表达，在受精卵中选择性表达父本还是母本的等位基因中起重要作用。

2. 染色质结构对基因转录激活的控制

真核细胞 DNA 是很长的分子，在核内以染色体形式存在，DNA 分子被组蛋白体包裹形成核小体，核小体再进一步压缩包裹成 30nm 丝状结构，最后可能还有更高级的结构形成异染色质。一个基因的转录必须允许激活转录的调控蛋白、转录因子和 RNA 聚合酶靠近并接触 DNA。染色质结构如何改变才能做到这一点？当 RNA 聚合酶沿着 DNA 模板移动延伸转录时，核小体又将如何变化？有些基因（占染色体相当大的一部分）在一些类型细胞中总是休眠的，核小体结构如何使它休眠？目前还不能清楚地回答这些问题，只有一点能肯定，就是基因转录的激活首先需要染色质结构的改变。

在电子显微镜下观察到两种染色质的丝状体：一种 10nm，另一种 30nm，10nm 染色质可见“珠-绳”结构，“珠”是圆柱形的组蛋白体，直径约 11nm，高 6nm，组蛋白体含 2 个 H2A·H2B 二聚体和 2 个 H3·H4 二聚体，成为八聚体结构。“绳”就是线状 DNA，在每个组蛋白体上螺旋缠绕 1.8 圈，每圈相当于约 80bp 长度，DNA 缠绕在组蛋白体上成为核小体，核小体是染色质的基本结构，每个核小体上缠绕的 DNA（146bp）加上两个核小体之间的部分共约为 200bp（约相当 67nm 长度），核小体结构不需要 H1 蛋白。10nm 的染色质是松弛和伸展的形式。

30nm 的丝状染色质结构，是 10nm 结构的进一步紧密压缩成螺旋，每圈螺旋含 6 个核小体，形成这种结构需要 H1 蛋白。30nm 结构是染色质和异染色质的基本结构。

基因表达需要染色质的改变，变得更加伸展和松弛，这方面最早的证据是用低浓度的 DNase I 消化，只在表达的基因区域发现对 DNaseI 的超敏感点，不表达的基因不存在酶的敏感点，表明在表达基因区域起始部分，DNA 是暴露的。在 SV40 染色体后期转录单位启动子上游一段 DNA 可以很好地被 DNase I、小球菌核酸酶及其他核酸酶（包括限制性内切核酸酶）切割，这段伸展的染色质长约 400bp，是自由的 DNA 片段，用电镜观察 SV40 染色体，也可看到这一段长约

120nm，相当于约 360bp 长度，电镜对 SV40 染色体的观察直接证实 DNase I 超敏感区是没有核小体的，当然这并不意味着在体内这段没有核小体的 DNA 上也没有其他蛋白。

染色质的结构控制基因转录，控制的方式是在基因的控制区改变结构，产生 DNaseI 超敏感区，超敏感区是一段长 200～400bp 的伸展的无核小体的暴露的 DNA。

关于产生超敏感点激活基因转录的机制，目前还有争论。生化研究的一些结果揭示出下面几个问题，比较转录的基因和不转录的基因的组蛋白区表明：①组蛋白 H1 与不转录区染色质结合比转录的染色质结合紧，H1 的五个丝氨酸磷酸化总是伴随着染色质凝聚和紧密堆积，有可能 H1 的去磷酸化使超敏感点的核小体堆积松弛而使染色质伸展。②在转录的染色质上组蛋白是高度乙酰化的。③H2B的磷酸化程度降低。④从果蝇到人类的许多种生物，转录染色质区域 H2A 的形式不同。⑤HMG14 和 HMG17 两种蛋白（high-mobility-group protein）只存在于转录的染色质。HMG 蛋白在进化上是高度保守的，它们与核小体结合可能在染色质结构改变允许基因转录上有重要作用。

除了上述组蛋白改变外，转录激活蛋白（包括转录因子）与超敏感区的 DNA 特异序列结合进一步改变染色质的结构当然是很重要的，这些蛋白的结合可能阻止组蛋白与 DNA 联系或取而代之。

一旦转录起始，RNA 聚合酶向前转录的移动中可以容易地通过核小体，不需要把组蛋白从 DNA 上驱逐出去，转录通过一个核小体时，两个 H2A·H2B 二聚体轮流暂时离开八聚体，八聚体的暂时改变是可逆的。图 10-27 是 Van Holde 等人 1992 年提出的 RNA 聚合酶转录通过核小体的模型。

3. 染色质的紧密堆积可防止基因转录

上面我们介绍了染色质松弛和伸展使基因激活，另一方面，染色质的紧密压缩堆积，可使调控蛋白和转录因子都不能靠近 DNA，而使基因组中相当大的一部分基因处于休眠状态，当然其中有些基因可以从休眠变成激活，另一些则不能转变。这些高度凝聚的染色质结构形式包括异染色质。

对酵母 S. cerevisae 的 *ADE*2 基因表达的研究表明，*ADE*2 在基因组中的位置不同，表达的效果就不同，这种现象称为位置效应（position effect），当 ADE2 蛋白不存在时，腺嘌呤生物合成途径被阻断；细胞累积红色色素，因而形成的酵母集落呈红色，正常集落呈白色，当 *ADE*2 基因处在正常位置时，所有的细胞中都表达 ADE2。而当处在端粒附近时，在绝大多数红色细胞中 *ADE*2 基因处在休眠状态。除 *ADE*2 不表达外，在染色体末端 10kb 的范围的基因都不表达。随着 *ADE*2 的位置离端粒越来越远，抑制的程度也逐渐减弱，位置效应是可以遗传的。端粒附近的基因表达受抑制，是由于核小体的堆积压缩十分紧密所致。

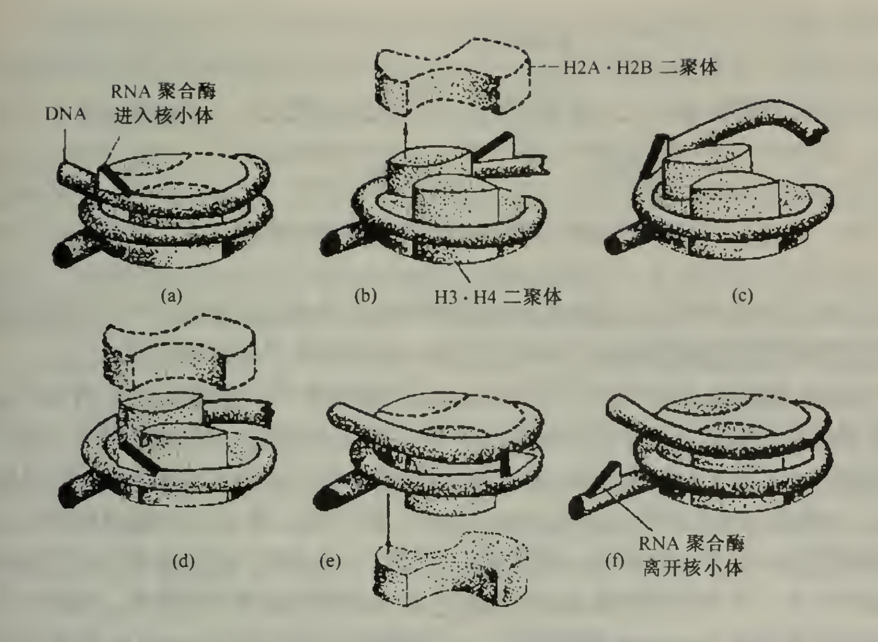


图 10-27　RNA 聚合酶具有在不引起组蛋白解体的情况下通过核小体的能力

4. 基因重排对基因表达的影响

(1) 同一个启动子连结不同的结构基因

一个结构基因通过移到其启动子处而被激活，在酵母菌的结合型控制中研究得最为清楚。在单体酵母中存在两种结合型，a 和 α 分别由 *HMa* 和 *HMα* 基因座位控制，a 型细胞可以有效地与 α 型细胞接合合并（merge）成 a/α 二倍体细胞，而同型细胞不能结合，即 a 型细胞之间不能结合，α 型细胞之间也不能结合。*HMa* 和 *HMα* 是非同源的基因簇，位于同一条染色体上，但彼此靠得不近，这个位置是非等位的，每个单倍体细胞中含有 *HMa* 和 *HMα* 两种基因簇，但一个单独细胞里只有其中一个基因是被表达的，例如，如果 *HMa* 表达，此细胞为 a 型细胞，在某种情况下，α 型细胞可以转变成 a 型细胞（反之亦然），这种现象叫做结合型转变，这种结合型转变现象只发生在含有 *HO* 基因的细胞里（HO, homothalic 的简写，即同型结合）。

结合型的转变依赖 *HO* 基因，*HO* 突变后，转变频率下降到 10^{-6}。

已有一些证据推导出图 10-28 中的结合型基因表达模型。

酵母的结合型暗盒调控模型中，*HMa* 和 *HMα* 本来处于无活性位置。*HMa* 和 *HMα* 基因可以移动到 Mat 座位，它们的移动需要 *HO* 基因产物，当 *HMα* 移动结合 *Mat* 座位时，HMα 表达，HMa 不表达，这个酵母细胞是 α 结合型。

模型中，Mat 座位邻接的是启动子。无论是 *HMa* 还是 *HMα* 插入到 Mat 时，

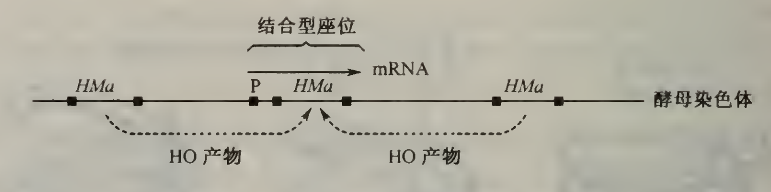


图 10-28　酵母结合型调节的暗盒式模型

就可以被启动转录表达，要说明一个重要的事实就是当发生插入时，原来的那个保持沉默的拷贝仍然留在原来的位置（这一点与转座子的转座相像)。

HMa 或 *HMα* 的一个拷贝移到 Mat 位并且原拷贝仍保留在原来的位点，这个过程类似于转化，而此基因移到染色体上固定的位置上（Mat）以后一般不再引起转位，而当 *HMa* 或 *HMα* 基因簇移到 Mat 时，这个位置原先已被 *HMα* 或 *HMa* 占据着（也就是说 mating-type locus 从未空过)。所以一个基因簇的进入必须伴随着另一个单位的移去。关于这一点已有一些证据，已经发现了这个移动的单位被丢失，这个过程类似于在溶源菌中的所谓原噬菌体替换现象。例如当 λ 噬菌体的溶源菌被 λimm434 超感染后，λ 原噬菌体被割离而 λimm434 被整合到宿主染色体上，通过类似的比较，假设可能 mating 座位上基因的互相转变是位点专一性的重组事件，它可能是被一种像 λ 的整合酶一样的酶介导的。目前 mating-type locus 已被克隆，在体外一种酶的确能使这个位点的双链 DNA 断开。

(2）免疫系统中编码顺序的连结

1）免疫球蛋白基因。把两种稍有不同的蛋白质（例如沙门氏菌的鞭毛抗原 H1 和 H2）注入到动物体内，就会使它产生两种专一性的抗体，据粗放的估计，抗体数可达百万种，蛋白质都是由基因编码的，那么怎样可能由基因来编码这许多免疫球蛋白分子呢?

20 世纪 40 年代化学家 L.Pauling 提出的指令学说似乎可以解答这一难题，按照这一学说，免疫球蛋白的一级结构只有一种或少数几种，所以相应的基因也只有少数几个，动物的淋巴细胞吸收一种抗原蛋白后细胞中的抗体分子便按照抗原的构形而改变它的构形，使两者之间可以发生专一性反应。但是以后许多生物化学研究结果认为蛋白质的二级或三级结构完全由一级结构所决定，也就是说从某一种一级结构只能得到某一种二级结构或三级结构，所以指令学说显然不能成立。因此必然有为数众多的基因来编码这些免疫球蛋白。

哺乳动物的总的基因数至多不过几万，所以很难想像每一淋巴细胞都能产生全部抗体。1957 年 F.M.Burnet 提出所谓克隆选择学说（即细胞选择学说)，认为每一种浆细胞只能产生一种或少数几种抗体。抗原进入体内后，刺激相应的淋巴细胞快速分裂并产生大量抗体。按照这一学说，抗原只能起着选择某一淋巴细胞克隆的作用。

一些简单的实验可以用来说明克隆选择学说基本上是成立的。实验一用两种不同的鞭毛抗原的细胞免疫兔子的淋巴组织，并把浆细胞分散悬浮，使每一滴液体中只有一个浆细胞，加入这种或那种鞭毛抗原的细菌，结果说明每种浆细胞所产生的抗体只能使一种抗原细菌不活动，说明每一个细胞只产生一种抗体。实验二用两种蛋白质免疫动物，同样取分散的浆细胞，另外再把这两种蛋白质抗原用两种不同的荧光染料结合成荧光抗原，其中一种抗原在紫外线下呈红色，另一种呈绿色，再把两种荧光抗原的混合物加到浆细胞悬液中，在紫外光下可以看到一些细胞呈红色，另一些呈绿色，没有一个细胞呈两种颜色。

克隆选择学说可以避免一个细胞中有成百万个免疫球蛋白基因这一概念上的困难，可是又怎么理解由一个受精卵产生如此众多的不同的免疫球蛋白基因呢？Burnet 认为体细胞突变可能是它的原因。不过这样高的体细胞突变率从没有在其他情况中发现。另外，免疫球蛋白由两个轻链和两个重链组成。每个链都有两个部分。一个称做可变区（V 区），另一个称为恒定区（C 区）。以轻链为例，它的可变区有 109 个氨基酸（1～109）（图 10-29），其中 68 个位置上的氨基酸多数发现过三种以上的变化；恒定区有 105 个氨基酸（110～214），其中只有三个位置曾发现过变化，到现在为止还没有发现过一个基因可以截然划分为突变率如此不同的两个部分，因此体细胞突变假说很难成立。W.G.Dreyer 和 G.C.Bennett 于是在 20 世纪 60 年代中提出另一假说，认为在浆细胞成熟过程中染色体发生重排。重排使原来编码两个链的基因连接在一起，于是就出现了一个可以区分易变区和恒定区的蛋白质分子。

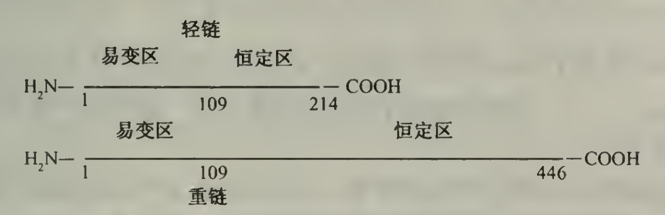


图 10-29　免疫球蛋白重链和轻链的可变区和恒定区

免疫遗传学的内容非常广泛，从真核生物的基因组结构角度来看，主要的问题是如果染色体重排假设可以成立的话，那么这些基因在基因组中以怎样的形式存在着？重排的机制又是什么？

免疫球蛋白的基本结构是一对轻链、一对重链结合成的 Y 形分子。人的免疫球蛋白的重链有五种类型即 γ、α、μ、δ 和 ε。其相应的 Ig 类型为 IgG、IgA、IgM、IgD 和 IgE。它们的恒定区的氨基酸顺序彼此间大约有 60% 差别。轻链只有两种类型（κ 和 λ），它们的恒定区的氨基酸顺序也有 60% 以上的差别，属于同一种类型的轻链或重链的恒定区几乎总是相同的，同一类型不同种的免疫球蛋

白的差别完全取决于易变区，免疫球蛋白的类型是以重链为标准来区分的（表10-3）。

表 10-3　人的免疫球蛋白种类

免疫球蛋白	IgG	IgA	IgM	IgD	IgE
重　链	γ	α	μ	δ	ε
轻　链	κ或λ	κ或λ	κ或λ	κ或λ	κ或λ
血清中的含量（%）	80	13	16	1	0.002

2) 染色体重排假设的证实。为了证实在个体发育过程中曾经发生有关免疫球蛋白基因的染色体重排，把胚胎的 DNA 和成体的 DNA 进行比较。成体 DNA 用骨髓瘤细胞 MOPC321（moure plasmacytoma321）作为代表，它只产生 κ 链。胚胎 DNA 用小鼠胚胎提取，实验过程如下：

①提取 MOPC321 的 DNA 和小鼠胚胎 DNA；②用限制酶 *Bam* H I 处理；③电泳区分大小不等的 DNA 片段；④从凝胶上洗下每一片段；⑤提取 MOPC321 的 κ 链 mRNA，另外取得它的 3′端的一半，两者都用碘（^{125}I）标记，制备探针；⑥把探针和洗下的 DNA 片段进行分子杂交测定，结果发现胚胎 DNA 能与完整 mRNA 杂交的片段有两个。它们的分子质量是 3.9×10^6Da 和 6.0×10^6Da。但是和 3′端一半进行分子杂交的只有分子质量为 6.0×10^6Da 这一片段，这一事实说明分子质量 3.9×10^6Da 这一段中不包括相当于 mRNA 分子的 3′端这一部分。

对于 MOPC321 DNA 来说，两者并无区别，而且进行分子杂交的片段只有 2.4×10^6Da 这一个，说明在胚胎 DNA 中原来不在一起的两个部分在成体中处在相近的位置上。

3) 免疫球蛋白基因多样性的来源。20 世纪 70 年代通过基因克隆和核苷酸顺序分析，一些免疫球蛋白基因的核苷酸顺序已被分析清楚。人的 κ 轻链的研究结果说明：①编码可变区的基因（variable， V）不是一个而是有大约 150 个，它们作为一个基因家族和间隔序列相间排列在一起；②每一个 V 基因有一个编码引导区的部分（L），它编码 17～20 个氨基酸，这一部分的功能是使免疫球蛋白分子通过细胞释放出来；③编码恒定区的基因（constant region，C）只有六个，它们组成另一家族；④V 家族和 C 家族中间还有五个连接序列（J），J 家族和 V 家族距离较远，C 家族则比较近。

在淋巴细胞发育过程中通过某种染色体重排过程 L/V 和 J 相连接，转录成包括 L、V、J、C 几个部分的 mRNA，再经过剪接成为成熟的 mRNA 分子，翻译产物上的引导区可能在穿过细胞过程中被去除（图 10-30）。

重链有五种，它们的基因的组织也更为复杂：作为易变区和恒定区的连接

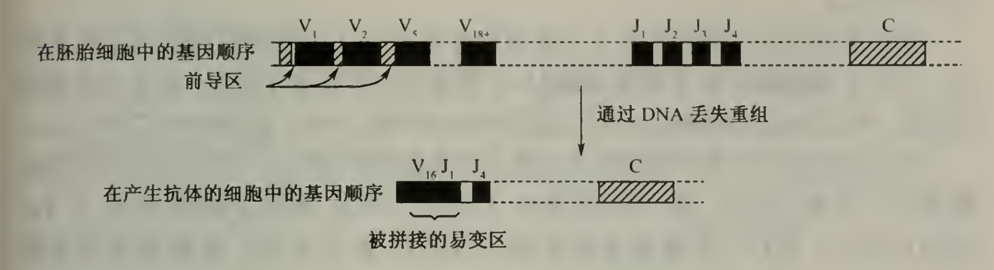


图 10-30 轻链可变区形成中的 DNA 拼接

部分除了连接序列 J 以外还有多样性序列 D；恒定区基因是八个；每个 C 基因中又分为几个区。

轻链的 V 基因有 150 个，J 序列有 5 个，每个 V 和每个 J 都不相同，所以 V 和 J 的连接有 150×5＝750 种。此外由于 V 和 J 的连接处并不十分准确，通过氨基酸顺序分析可以看到 V 和 J 的交接处的氨基酸有较多的变化，假设由于交接处重组位置的细微差别带来的变化有 10 种，那么轻链的种类就有 150×5×10＝7500 种。

对于重链来讲，估计人的 V 基因有 80 个，J 序列有 6 个，D 有 50 个，所以通过染色体重排可以得 80×6×50＝24 000 种组合，由于重组带来的变化估计有 100 个，因此总的变化 2 400 000。再考虑轻链和重链的随机组合，免疫球蛋白分子的种类将达 18 亿之多。

4）染色体重排机制。为了研究染色体重排机制，必须对胚胎的免疫球蛋白基因和成体的免疫球蛋白基因进行比较，在这项研究中，能产生单一种类的免疫球蛋白分子的骨髓瘤细胞株仍然为最合适的材料。

对小鼠的 κ 轻链 DNA 的酶切图谱分析和核苷酸顺序分析结果说明 J 基因有 5 个，它们的排列方式是：

（310）　　（246）　　（218）　　（299）

……J1……………J2…………J3…………J4…………J5

（38）　　（38）　　（38）　　（38）　　（36）

括弧中的数字表示核苷酸的数目，5 个 Jκ 序列的核苷酸顺序中大约 84％都是相同的，不但如此，J 序列的两旁核苷酸顺序也是非常保守的。

J 序列的这种核苷酸排列方式不限于 Jκ，也包括 Jλ、JH，J 的 3′端的非编码区，其共同顺序是：3′………GTTTTTGT-（21～23base）-CACTGTG……5′。V 序列的核苷酸顺序分析的结果表明 V 的 3′端的非编码区的共同顺序是：3′………CACAGTG-（11～12base）-A-CATAAAC-5′，这两个序列显然有反向重复性质，看来 V/J 连接处的核苷酸顺序有下列通式：

V-CACAGTG-（11 ～ 12base）-ACATAAACCTTTTTGT-（21 ～ 22base）-

CACTGTG-J

基于上述发现，现在认为 V-J 重排是通过 V 和 J 中间的序列的切除而实现的。在这个过程中 V 的 5′端和 J 3′端反向重复序列形成茎环结构，使它们的编码区靠拢，然后通过某些链的作用而切除茎环结构并使 V 和 J 连接起来（图 10-31）。

这一连接过程中剪切处稍有移动便会带来前面所说的 V-J 交接处的多种氨基酸变化，如图 10-32，如 V-J 交接按 1 进行，那么交接处的氨基酸是 Trp (TGG)；按 2 进行，氨基酸是精氨酸（CGG)；按 3 进行，氨基酸是脯氨酸 (CCG)；按 4 进行，氨基酸是脯氨酸（CCC)。

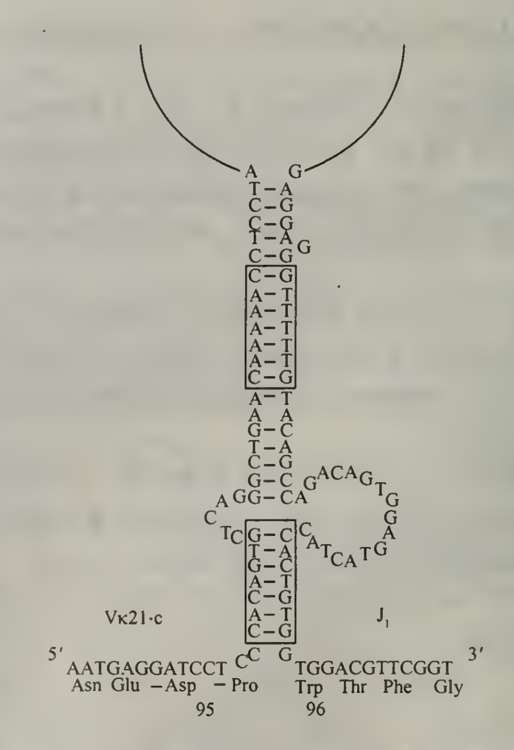


图 10-31　免疫球蛋白基因拼接处的碱基顺序

值得注意的是，图 10-30 中 V-J 基因的拼接是 DNA 的拼接，而不是 RNA 的拼接。V、J 基因 DNA 之间的拼接是基因重排的一种方式，是免疫球蛋白基因重排的第一步。转录后 mRNA 还要经加工拼接最后翻译出免疫球蛋白（图10-33)。从图中可以看出：①J 与 C 基因之间的间隔区仍然存在。②有活性的 L 链只来源于一个 V 和一个 J，但拼接后的 DNA 中仍有许多 V 和 J 基因。获得正确的氨基酸顺序要通过 RNA 的最后拼接得到成熟的 mRNA。

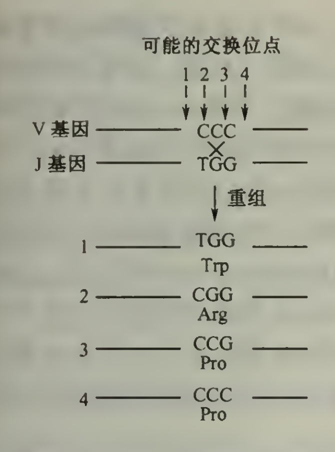


图 10-32　V-J 交接处的可能位置

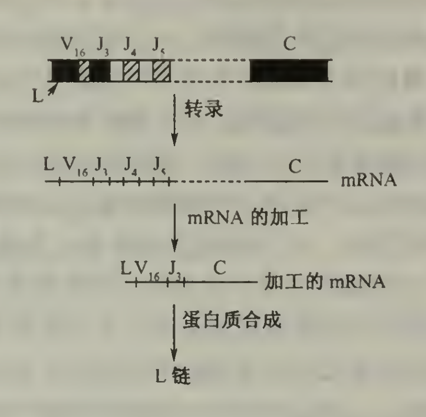


图 10-33　IgG 分子 L 链的产生

五、转录后的控制

真核基因转录产生 hnRNA，通过 5′和 3′末端修饰、剪接、编辑等一系列加工过程才成为成熟的 mRNA，然后被运送到细胞质的特定区域才能翻译，这些过程对基因表达水平都会产生影响。

1．转录延伸的弱化

转录起始之后，每个基因转录并不一定以相同的速度转录到终点，其中一些转录可由弱化控制中途停顿甚至提前终止而使转录流产，转录弱化在原核中资料较多，但在真核中同样存在，不过弱化机制有所不同，在真核中转录的弱化程度（RNA 多聚酶是否能通过下游的特异弱化位点）可能取决于蛋白质（如 *hsp* 基因的转录），即包括在启动子上装配的蛋白质或结合在转录起始点下游的蛋白质。这些蛋白质在不同类型细胞中不同，细胞能控制不同基因的弱化程度。

2．RNA 的剪接加工和编辑使同一基因转录产生不同的蛋白

剪接过程中据估计有半数以上的 RNA 被完全降解，有半数可产生不同形式的剪接产物，使基因可产生不同的蛋白质。剪接的改变可以是组成型的，这种情况下拼接改变是有限的，每个转录单位只产生少数几种蛋白质。但在许多情况下，改变剪接是受控制的而不是组成型的，在这种情况下，某一基因转录物能否被正确剪接而最终产生有功能的蛋白质受到细胞类型的控制。例如果蝇中的一种催化 P 因子转座的转座酶，只在生殖细胞中产生有功能的蛋白，在果蝇的体细胞中均产生无功能的蛋白，原因是一个内含子序列只有在生殖细胞中被剪接去除。

剪接加工对基因表达也可有正和负控制之分，上述 P 因子转座酶在果蝇体细胞中的加工是负控制，在生殖细胞中的加工是正控制。另外，根据细胞的需要，剪接加工控制在不同类型细胞中产生不同的蛋白质，例如 *Src* 原癌基因编码的酪氨酸蛋白质激酶，在大多数类型细胞中产生 533 个氨基酸组成的蛋白，但在神经细胞中却产生 539 个氨基酸的蛋白质，这是因为一个小的内含子 A 在神经细胞的加工中保留，神经细胞中的这种酪氨酸激酶多了一个磷酸化位点，具更高的特异活性。又如果蝇发育中有三个基因——*sxl*、*tra*、*dsx* 决定性别，是在 RNA 的剪接加工水平上控制，剪接方式不同，*sxl* 和 *tra* 在雄性发育中产生无功能的蛋白，在雌性发育中产生有功能的蛋白，dsx RNA 剪接不同，产生 C 端不同的 Dsx 蛋白，在雄性个体发育中的 Dsx 蛋白 C 端 150 个氨基酸是雄性特异的，雌性的 Dsx 蛋白 C 端 30 个氨基酸是雌性特异的。

剪接的正负控制决定剪接的形式，控制的机制是在负控制中，RNA 的剪接位点上有抑制蛋白结合，使本来可以发生的剪接不能进行，在正控制中，剪接机器本不能在某一位点剪接，但由于剪接的激活蛋白在这位点附近结合而使剪接得以在这点上进行，不同类型细胞根据自己对某一基因产物的需要产生剪接加工的激活蛋白或抑制蛋白。

在一些 RNA 的加工中产生两种以上的功能蛋白，例如 HCMV 的 IEII86 和 IEII55 都是由同一 hnRNA 产生，两种的比例相对稳定，如何控制不同形式加工的比例还不清楚，估计是由剪接加工的抑制或激活蛋白的量控制，也可能是由两者的比例决定。

RNA 编辑在 RNA 的多处增加 U，可造成读框改变，出现终止码，改变氨基酸等而产生不同蛋白。

3．RNA 的切割和加 polyA 也能改变基因编码的蛋白质

真核 mRNA 的 3′端不是由 RNA 多聚酶 II 的合成终止决定的，而是在转录延伸过程中由另外的因子催化切割决定的。在一般情况下切割位置在 polyA 信号序列 AAUAAA 的下游约 30 个核苷酸左右，然后再在末端加上 polyA，但细胞有能力控制在 RNA 的不同点切开。研究得较清楚的例子是 B 淋巴细胞发育过程中合成膜结合抗体和分泌性抗体，B 淋巴细胞早期接触抗原后受刺激合成的抗体的 C 端是结合在质膜上的，其余部分暴露在细胞表面作为抗原受体，当第二次接触这种抗原，表面的受体与抗原形成复合物刺激 B 细胞分裂和合成分泌性抗体。膜结合型和分泌型抗体只是 C 端的氨基酸序列不同，前者含较长的亲脂性（疏水的）氨基酸序列，而后者 C 端含短的亲水的氨基酸序列。

4．RNA 从核中运送到细胞质的过程也是受控制的

在合成的 RNA 总量中，约有 1/12 的 RNA 被运出核到细胞质中，其中在核

内被完全降解的约 50%，另一些被留下的 RNA 可能含一些序列，这些序列使它们决不会成为 mRNA，另有一些代表潜在的 mRNA，它在某些类型细胞中是有功能的，而在另一些类型细胞中则不能被运出核。

虽然 RNA 运出核的控制机制目前不了解，运出核的控制形式的证据也很少，但至少有几点可以说明 RNA 运出核外是受控制的：①RNA 通过核膜孔运出的过程是一种激活过程；②对大多数 mRNA 需要有 5′帽子结构和 3′polyA 结构；③mRNA 必须与所有的剪接体成分完全脱离后才能被运出核。

mRNA 通过核膜孔从核中运出，它被送到细胞质中的位置也有特异性。例如，如果 mRNA 编码的蛋白质是分泌到胞外的，它被直接运到内质网，连同与之结合的核糖体和新生的肽链一起被运输，即运输与翻译同时进行，在内质网膜上继续完成肽链的合成。而另一些 mRNA 则可能被运到细胞浆液中，由细胞浆中游离的核糖体进行翻译。在细胞质中的运送和定位，前一种情况是由新生肽链 N 末端的信号序列（信号肽）决定的，这段序列刚被核糖体合成，就立即被细胞的一种“蛋白质截短成分”所识别，这种成分指导将 mRNA-核糖体-新生肽复合物运到内质网膜。在另一些情况下，运送和定位在翻译之前由 mRNA 本身的核苷酸序列决定，这段信号序列典型的是在翻译终止码和 polyA 开始之前的 3′不翻译区域。

六、真核蛋白质合成的控制

在蛋白质合成水平上的控制，是基因表达控制的重要环节，同一细胞中同时出现的不同 mRNA，即使 mRNA 数目接近相等，产生的蛋白质的多少也可以差别很大。这主要取决于翻译的速率和 mRNA 的寿命。

1. mRNA5′前导序列对翻译水平的影响

真核 mRNA 翻译起始，对绝大多数的 mRNA 来说，是核糖体小亚基从 5′帽子沿着 mRNA 移动寻找 AUG 密码，一般遇到第一个 AUG 就与核糖体大亚基装配后开始翻译。但也有的漏过第一个 AUG，从第二个或第三个 AUG 开始翻译。只有极少的真核细胞和病毒的 mRNA 翻译时核糖体小亚基是从中间进入的，这些 mRNA 的内部有一核糖体进入的位点。另外，在翻译过程中如果核糖体遇到弱化序列或终止信号，真核核糖体会立即拆散，并且不会再在此处重新装配使翻译重新开始。

在疱疹病毒（HSV）感染的细胞中，TK（胸腺嘧啶激酶）和 pol（DNA 多聚酶）两种病毒特异的 mRNA 在细胞中同时出现并且数量（分子数）接近相等，但翻译产生的 pol 蛋白还不到 TK 蛋白的 1/20，电镜观察发现 pol mRNA 上很少有核糖体结合。HCMV 的 DNA 多聚酶的 mRNA 的 5′前导序列与 HSV 的 pol 的

mRNA 十分相似，是一段长的富含 GC 的序列，这段序列可以与紧接 AUG 下游的序列形成含多个 GC 的茎环结构或发夹结构，AUG 处于环中，这样核糖体小亚基很难移动到 AUG 位置。昆虫杆状病毒包含体蛋白和一些植物病毒衣壳蛋白的 mRNA 翻译效率非常高，它们的 5′前导序列富含 A 和 T，如果用植物病毒 *AMVR*4（衣壳蛋白基因）的 mRNA 的前导序列代替 HCMV pol 的 mRNA 的前导序列，pol 的 mRNA 的翻译水平就会显著提高。有人用 *AMVR*4 的前导序列取代一种肿瘤抑制蛋白 Rb mRNA 的前导序列，体外翻译水平提高几十倍；铁蛋白 mRNA 的近 5′端也有茎环结构，一种翻译抑制蛋白叫顺乌头酸酶（aconitase）结合在这茎环结构区域，完全阻止铁蛋白 mRNA 的翻译。当有铁离子时，Fe 离子与这抑制蛋白结合使它与 mRNA 分离后，mRNA 才能被翻译产生铁蛋白。另有些在翻译水平上进行反馈抑制的蛋白质，当这种蛋白质达到一定浓度时，部分蛋白质分子结合在编码它的 mRNA 5′端阻止翻译。

上述 5′前导序列形成茎环结构降低翻译水平或抑制蛋白结合 5′端阻止mRNA 的翻译，都是阻止翻译的起始，即阻止核糖体小亚基向 AUG 移动而达到降低或封闭 mRNA 的翻译。另有一种战略同样能降低从正确的 AUG 开始的翻译，就是在功能蛋白的编码区下游另有一个 AUG，核糖体小亚基必须漏掉这个 AUG，从第二个 AUG 开始翻译才能获得有功能的蛋白，而核糖体常从第一个 AUG 开始翻译，产生 10～20 个氨基酸肽后即终止。这样，第一个 AUG 对mRNA的正确翻译起干扰或弱化作用。当然，有些 mRNA 从不同的 AUG 翻译可产生几种不同的功能蛋白，这种现象常见于病毒，可使病毒利用有限的基因组编码更多的蛋白。

2．翻译起始因子的磷酸化可抑制或选择性地加强蛋白质的合成

真核细胞在被病毒感染、热休克、去除某些生长因子、进入细胞周期的 M 期等情况下都抑制大部分蛋白质的合成，而少数蛋白质的合成反而加强；这种控制中，许多与翻译起始因子 eIF-2 的变化有关，eIF-2 被特异性的蛋白质激酶磷酸化后，降低大多数蛋白质合成的速率。

有正常功能的 eIF-2 与 GTP 形成复合物，介导甲硫氨酰 tRNA 与 40S 核糖体亚基结合，然后这种复合物进入 mRNA 5′端并移动到 AUG，eIF-2-GTP 转变成无活性的 eIF-GDP 从复合物中释放出来，60S 的核糖体亚基这时和 40S 小亚基装配后，蛋白质进行合成。eIF-2 与 GDP 结合很紧，必须使 GDP 从 eIF-2 中释放出来，才能使 GTP 重新与 eIF-2 结合，eIF-2 才能重新被利用。这个 GDP 和 GTP 变换过程需要一种叫 eIF-2B 的蛋白质。如果 eIF-2 磷酸化，就能抑制 eIF-2 重新利用，因为磷酸化的 eIF-2-GDP 和 eIF-2B 的结合非常牢，防止 GDP 和 GTP 交换，在细胞中通常 eIF-2 的数量比 eIF-2B 多，即使一部分 eIF-2 磷酸化，就能结合几乎全部的 eIF-2，使未磷酸化的 eIF-2 也难以被重新利用，这样就使蛋白质合成大大减少。

从上面所述中，人们很自然地会想到，当 eIF-2 磷酸化时，所有的 mRNA 的翻译将会同时降低，但与这种想法相反，eIF-2 磷酸化只是选择性地降低一些 mRNA 的翻译，对另一些 mRNA 的翻译反而加强。酵母在缺乏氨基酸的培养基中，eIF-2 磷酸化可加强与合成氨基酸有关的蛋白质的合成而降低其他所有蛋白质的合成，使酵母适应无氨基酸的饥饿环境。调节 eIF-2 的水平对哺乳动物细胞进入 G0 期是很重要的，G0 期蛋白质合成的速率只有增殖细胞的五分之一。

3．mRNA 3′端的结构对翻译速率和 mRNA 寿命的影响

（1）polyA 的影响

一般真核 mRNA 的 3′末端有一长的 polyA，这段长的 polyA 除影响 mRNA 的稳定性外还影响 mRNA 的翻译，因此 polyA 从两方面影响蛋白质合成的水平。polyA 长度在 30～200 个腺嘌呤核苷酸范围，一些 mRNA 的 polyA 长度是受控制的，polyA 对于 mRNA 的稳定性是需要的。细胞可以对不同的 mRNA 的 polyA 选择性加长或快速截短、去除，那些被去除 polyA 的 mRNA 就很快被降解，那些被加长 polyA 的 mRNA 保持稳定，寿命长，可多次翻译。

polyA 除与 mRNA 的稳定即寿命有关之外，还可影响翻译，在成熟的卵中，通过加长特异的 mRNA 的 polyA 的手段控制基因表达就是一个引人注目的例子。在巨大的卵细胞中，许多正常的降解 mRNA 的途径是不奏效的，因此细胞储存大量的 mRNA 以备受精，这些储存的 mRNA 3′末端的 polyA 只有 10～30 核苷酸长度，这种形式的 mRNA 是不翻译的。受精后细胞需要这些 mRNA 编码的蛋白质，就在需要翻译的那些 mRNA 的 3′端再增加 polyA，逐渐刺激这些 mRNA 起始翻译。

加长 polyA 刺激 mRNA 的翻译的机制，一是通过 eIF4G 与 PablP 作用促进 40S 亚基与 mRNA 结合，二是 polyA 结合蛋白催化核糖体大亚基与结合在 5′端的核糖体小亚基的装配，如图 10-34 所示。

（2）3′端的序列对 mRNA 翻译的影响

组蛋白的 mRNA 的 3′末端没有 polyA，它们的 mRNA 的稳定性受 3′端的信号序列控制。当细胞进入 S 期，DNA 合成需要大量的组蛋白，这时组蛋白 mRNA的半衰期约 1h，在不合成 DNA 时，半衰期只数分钟，如果用药物在 S 期抑制 DNA 合成，组蛋白 mRNA 立即变得不稳定而很快降解。

组蛋白 mRNA 的 3′端存一段短的茎环结构代替 polyA，这一末端是在 RNA 聚合酶在合成后通过特异性切割而产生的，切割过程需要存在于一种核蛋白中的小 RNA 与切割区域进行碱基配对，如果用 DNA 重组技术把这段茎环结构 DNA 转移到别的 mRNA 3′端，别的 mRNA 也会随着 DNA 合成停止而变得不稳定。这说明 mRNA 的降解速率受这 3′信号的影响，而且降解是从 3′末端开始的。如果在组蛋白 mRNA 编码区中间插入一个翻译终止密码，这种 mRNA 是稳定的，

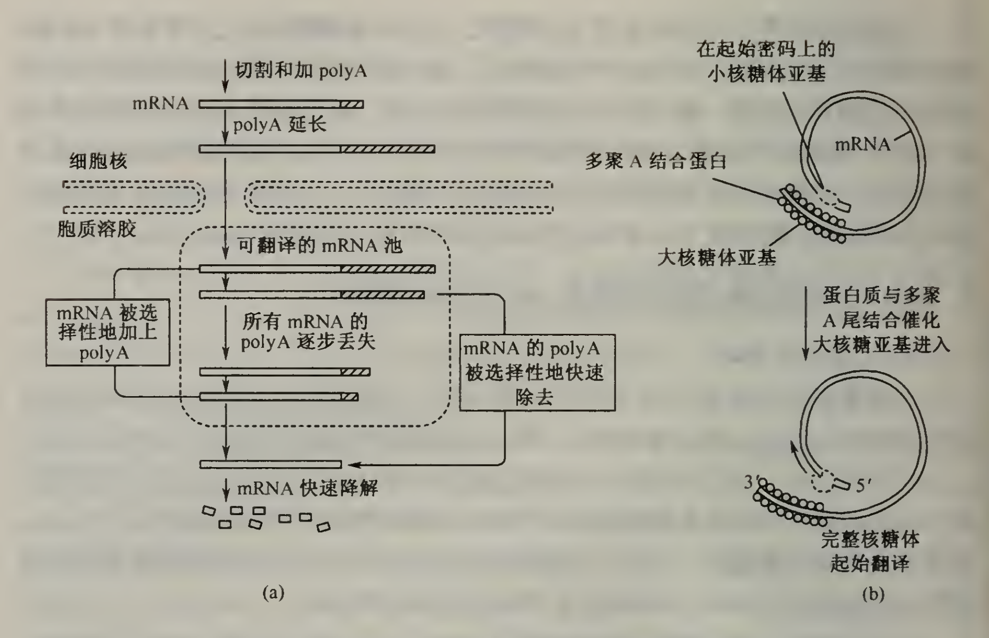


图 10-34　mRNA 3′端多聚 A 尾巴对基因表达的调控

即使细胞 DNA 合成停止也不会很快被降解，这说明降解 mRNA 的核酸酶是结合在核糖体上的，并且降解是从 3′端开始的，中间插入的翻译终止密码使核糖体提前与 mRNA 分离而不能靠近 3′末端。这种假设能解释为什么大多数不稳定的 mRNA 在用蛋白质合成抑制物环己亚胺处理细胞时变得稳定，这种与翻译偶联的 mRNA 降解机制可以保证 mRNA 在降解之前至少被翻译一次。

4．真核生物 mRNA 翻译起始的控制

在原核生物中，翻译的起始是通过核糖体小亚基上 16S rRNA 与 mRNA 上 SD 序列配对，mRNA 将核糖体召集过来。

而在真核生物中，翻译的起始依靠蛋白质-蛋白质、蛋白质-RNA 之间相互作用过程，这些过程是受控制的。以下为几种翻译起始因子及其作用。

eIF3：40S 小亚基联系因子，至少由 8 个亚基组成（哺乳动物），是与 mRNA联系的起始因子，与 eIF4F 相互作用。

eIF4F：所有真核细胞中的 eIF4F 都由帽子结合蛋白 eIF4E 和大亚基eIF4G 这两个核心亚基组成。

eIF4G：eIF4G 的 N 端与 eIF4E 相互作用，C 端与 eIF3 接触，eIF4G 上从 N 端到 C 端有 Pab1（polyA 结合蛋白）、eIF4E、RNA/eIF3 和 eIF4A 等不同因子的结合区，在 eIF4E 结合区和 RNA/eIF3 结合区之间有病毒蛋白酶切割位点。

eIF4A：在高等真核生物中，eIF4A 对 mRNA 帽子刺激的翻译起始中的 40S

小亚基的召集是必需的，eIF4A 与 eIF4G 结合，eIF4A 表现出依赖与 ATP 的 RNA 解旋酶活性，这种活性可由 RNA 结合蛋白 eIF4B 激活，但 eIF4A 与 eIF4B 不会发生直接的相互作用。eIF4A 和 eIF4B 对 40S 亚基召集的作用仍不十分清楚，可能是与 40S 亚基装配中的 RNA 解旋有关，提纯时，eIF4A 不与 eIF4E 一起共纯化，说明 eIF4A 不参与 eIF4F 复合物。

eIF4F 的装配：细胞合成一种抑制蛋白，可调控 eIF4F 的装配，被称为 eIF4E 结合蛋白（4E-BPS），这种蛋白有一段与 eIF4G N 端类似的氨基酸序列（XYDRXFLXXX，eIF4G 中的与 eIF4E 结合区的序列：KKRYDREFLLGF），当 4E-BPS 磷酸化时不能与 eIF4E 结合，但它去磷酸化后成为 eIF4G 与 eIF4E 结合的竞争抑制物。腺病毒（ADV）感染细胞可诱导 4E-BPS 去磷酸化，加速 eIF4E 与 4E-BPS 结合，阻止 eIF4E- eIF4G 的形成，从而抑制细胞 mRNA 的翻译。4E-BP1 是 4E-BPS 中研究最充分的成员，生长因子可诱导 4E-BPS 磷酸化（Ser_{64}磷酸化，由 MAP-有丝分裂激活蛋白激酶引起）。eIF4E 本身的磷酸化加速它与 mRNA5′帽子结合及 eIF4E 与 eIF4G 相互作用，eIF4E 的磷酸化（Ser_{209}）在细胞内是受调控的，但与 eIF4E 磷酸化有关的酶还没有找到。当 ADV 感染时，细胞内未磷酸化的 eIF4E 积累，降低细胞 mRNA 的翻译，而病毒的 mRNA 照常高效翻译，是由于病毒 mRNA 的分为三部分的前导序列（tripartite leader sequence）在低浓度的磷酸化 eIF4E 条件下可被优先选择翻译，它是病毒感染的又一种战略。polyA 结合蛋白（PABP），对 eIF4F 的装配有促进作用，研究表明，PAB1 刺激使翻译水平提高 2 倍以上，polyA 结合蛋白与 mRNA 的 polyA 结合及其功能又受 polyA 上游的 3′UTR 影响，3′UTR 可激活或抑制它与 polyA 结合及其功能，PABP 与 eIF4G 结合并互相作用，这种互相作用又促进 4E 与 4G 之间的结合。共同使 mRNA 成环，促进 40S 亚基与 mRNA 的结合及 60S 亚基的装配。

除 eIF4E-eIF4G 对翻译调控外，还可以通过 mRNA 上紧接帽子的（Cap-proximal）一段序列来控制［5′末端寡聚嘧啶序列（5′TOP)］，eIF4E-eIF4G、翻译延伸因子 1A 和翻译延伸因子 2 以及 polyA 结合蛋白可与 mRNA 调控序列 5′TOP（5′TOP 是紧接帽子的一段序列）相互作用。例如，在铁蛋白 mRNA 中，称为离子应答因子（IRE）的一段序列就位于 5′TOP 内，这段序列与离子调控蛋白（IRP）有很高的亲和力，IRP 与 IRE 结合后可以阻止小亚基的装配。

IRES 因子及核糖体 40S 亚基召集（中间进入）：正链 RNA 病毒的 RNA 5′NCRS 长达 300～1200nt，含有多个“AUG”和一段称为“中间核糖体进入序列”（IRES）及二级结构，也能在细胞内有效翻译。正链 RNA 病毒感染抑制宿主细胞的帽子刺激的翻译。而病毒的 mRNA 翻译由 IRES 介导，核糖体从 5′非编码区内部进入而不是从 5′末端进入，这种翻译机制也称“中间进入”机制。小 RNA 病毒的 mRNA 5′NCR 可分成三类：Typ I（poliovirus、enter virus and rhinovirus）Typ II（ardio-and aphtovirus，如 EMCV）Typ III（HAV，甲型肝炎

病毒)。

在5′NCR的近3′端有嘧啶丰富的保守结构（一级及二级结构），不同类型的差别在于这段序列与起始密码AUG的距离。在typ I中这段序列在起始码AUG上游50～100nt，而在typ II和typIII中AUG紧接这段序列的3′端边界，预料这段序列与核糖体小亚基中的18S rRNA互补，核糖体从这里进入，这段序列的一级结构和二级结构对核糖体进入都是必须的。

Typ I IRES的翻译效率在无细胞系统中很差，在不同的细胞内翻译效率差别很大，在eIF4G被病毒蛋白酶切开情况下，翻译效率显著提高。

Typ I mRNA通过这段序列与18S rRNA序列互补结合核糖体亚基，然后核糖体沿着mRNA向下游移动扫描出AUG。

40S亚基与Typ I或Typ III mRNA在通过互补结合后，还要重新放在AUG上，重新放置过程需起始因子作用，对HCV RNA翻译研究表明，包括起始码AUG在内的一小段编码区RNA序列对于重新放置是必需的。

已发现有几种细胞的mRNA也含有这类5′NCR，如免疫球蛋白的重链结合蛋白Bip的mRNA，成纤维细胞生长因子2蛋白、人eIF4G的mRNA等，细胞的这些mRNA的翻译起始也采用核糖体中间进入机制。

5. 翻译延伸过程对蛋白质合成的影响

在翻译起始之后，核糖体并非一定能从起始密码顺利移动到终止密码，mRNA编码区中间也形成一些二级结构，主要是小的茎环结构成为核糖体沿mRNA移动的障碍，使翻译速度下降甚至中途停止，部分核糖体可越过障碍翻译到终点。

另外有些mRNA（常见于反转病毒）在翻译延伸中发生“移框翻译”(translational frame shifting)，改变翻译的产物。当翻译延伸时发生移框翻译，因而清除了终止密码使翻译继续进行，使原来的编码改变而重新编码，这种现象称为翻译的重新编码（translational recoding)。移框翻译可以在翻译时把阅读框向前（+1）或向后（-1）移动一个核苷酸。移框翻译需要mRNA上发生移框的位置下游的特异序列，称为重新编码信号。

HIV Gag蛋白合成中，约90%的核糖体在翻译中没有发生阅读框改变，约10%的核糖体在翻译到UUU UUA GGG时（Phe-Leu-Gly)，在亮氨酸加入肽链后向后移动一个核苷酸，使下一个甘氨酸读框（GGG）变成精氨酸（AGG)，从而使以后的阅读框在翻译中都向后移动一个核苷酸，于是产生gag-pol（反转录酶）融合蛋白。

另有一类非常罕见的重新编码的形式，是在翻译延伸至终止码UGA时，受到此终止码附近信号影响，翻译不会终止，而是渗入一个含Se原子代替S原子的修饰的半胱氨酸，翻译继续下去，产生另一种蛋白。Se原子对一些酶的活性是必需的。

6. 微小的干扰 RNA 对 mRNA 翻译的负控制

在动植物细胞中普遍存在微小的干扰 RNA（siRNA），大小为 21～23 核苷酸，可以与 mRNA 互补结合，造成 mRNA 不能翻译和降解。这些微小的干扰 RNA 对 mRNA 的翻译起负控制作用。

第十一章　细胞信号调控

前几章论述了基因表达调控，可以说是基因水平上的分子生物学。而细胞信号则是细胞水平和整体水平的分子生物学。

在多细胞组成的生物中，尤其是高等动物体中，有多种组织、器官，每种组织和器官又含许许多多细胞。然而高等生物的个体是高度协调统一的一个整体，对外界的刺激能做出迅速而准确的反应。例如一个人受到惊吓时，心跳加速、脸色苍白、全身冒冷汗；运动员在做体操表演时全身的运动是那样协调；人体发育到一定生理阶段具有男人或女人的性特征等。神经、情绪、思维、运动以及各种生理、病理状态下，机体的不同部位都做出不同的协调的反应。这就需要在细胞之间进行有效的信息联络。细胞接受了这些信息，能正确地决定自己应该做什么，是死亡、是生存、是增生，还是进行一些特别的反应（例如改变细胞内基因表达的模式）。这就是细胞信号（cell signaling）。机体通过细胞信号对体内的许多细胞进行“社会式”控制。一旦信号控制失灵，就造成机体失调，甚至恶性增生和发生肿瘤，导致机体死亡。细胞信号联络即使在单细胞生物中也存在，例如酵母和大肠杆菌的接合交配是通过释放出一种称为交配因子（mating factor）的肽向周围细胞发出信号的。不过，高等动物的细胞信号系统更加复杂和精细微妙。

第一节　细胞信号的一般概念

一、信号分子和信号受体

显然，细胞信号最少包括两方面因素：①细胞外的信号分子；②信号受体。信号分子有好几百种，包括蛋白质、小肽、氨基酸、核酸、固醇类、脂肪酸类似物质、类维生素 A（retinoids），甚至小分子可溶性气体如 CO 和 NO。绝大多数信号分子由发出信号的细胞通过外排作用（exocytosis）或扩散作用穿过质膜（小分子信号）而分泌到胞外。另有一些信号分子并不分泌到胞外，而是紧紧地结合在发出信号的细胞外表面，这些信号只能影响与发信号细胞接触的细胞。信号分子的作用浓度很低（小于 10^{-8}mol/L），而且浓度可在短时间内变化。

受体又分细胞表面受体和细胞内受体。受体是靶细胞接受信号或对信号做出初步反应的一类特殊的蛋白质。细胞表面受体是一类跨膜蛋白质。受体识别信号分子能以很高的亲和力与信号分子结合（亲和常数 $K_a \geq 10^8$mol/L）成为激活状

态，引起靶细胞内一连串反应而改变细胞的行为。另一些细胞内受体，是催化反应的酶（如鸟苷酸环化酶）或基因转录的调控蛋白，是与分子质量足够小的疏水性信号分子结合的，因为疏水（亲脂）的很小的信号分子，如NO、CO和一些激素可以直接扩散穿过质膜和核膜进入靶细胞内激活细胞内受体蛋白，由被激活的受体蛋白直接催化反应或控制基因转录。

信号根据来源、作用范围和送达的形式可以分为三种形式。

1. 近分泌信号

近分泌信号（paracrine）由细胞分泌后扩散到它周围的细胞，因此它的作用只是局部的，不能远距离输送，也不允许扩散太远。当这种信号分泌后会立即被附近的细胞接受，或很快被细胞外的酶降解或很快被氧化（如NO、CO），或被胞外的基质固定不能再扩散。

2. 内分泌信号

内分泌信号（endocrine）由各种腺体或特殊组织产生，如各种激素，这种形式的信号由血液循环系统（植物则通过体液，如液汁）输送到全身，受血液稀释，浓度很低。

3. 神经信号

神经信号（synaptic）是由神经细胞或神经元受外界环境信号或别的神经细胞发出的信号激活而发出的电脉冲，电脉冲沿着神经轴突高速传递（每秒可达100m）直至轴突的末端，刺激末端分泌一种称为“神经传感器”（neurotransmitter）的化学信号，这些化学信号被快速（距离不到100nm，时间少于千分之一秒）而特异地传递到与神经轴突接触的靶细胞。关于神经信号的传递过程，这里不作具体介绍。

还有一种叫“自家分泌信号”（autocrine signaling）。上面介绍的信号是影响远或近距离的别的细胞的，而自家分泌信号则是影响周围与自己同类型的细胞，包括分泌信号的细胞本身在内，自家分泌信号在发育分化过程中决定一组细胞的分化是很有意义的。

二、细胞对信号的反应

每个细胞周围有许多各种各样的信号分子，可以说是处在信号分子的包围之中。细胞可以通过自己的表面受体和胞内受体获得信号，也可以通过连接相邻细胞的通道而共享一种信号。几百种信号分子，除可单独作用外，又可以进行组合可能产生出几百万种组合方式，以组合方式起作用，就像由4种核苷酸可以组合

成 60 多种氨基酸密码一样。那么细胞接受哪些信号，做出何种反应，显然也是千差万别的。

1. 每个细胞对某种特别的组合信号的反应是由程序决定的

在发育生长和分化的过程中，每个细胞内都形成一套独特的程序，就像我们在计算机中储存程序，当我们用指头按动某一个键，计算机就按程序进行工作，当信号与受体结合时，细胞就按预定程序规定做出各种反应。

2. 细胞对不同的信号反应不同

细胞在发育过程中获得了自己的特性，对信号的接受和反应是有选择性的，比如对一组信号按程序作出的反应是增生，对另一组信号的反应是行使某种特别的功能等。在培养的一瓶细胞中，有的细胞可以对一组信号做出按程序自杀的反应（即 programmed of cell death，也叫 apoptosis）。

3. 不同细胞对同一信号的反应可以不同

在不同类型的细胞中接收同一种信号的受体可以不相同，对信号意义的解释也不相同，因而做出的反应不同。在许多情况下，即使在不同类型细胞中接受同一种信号的受体完全相同，也可以对信号做出不同的反应，这是因为在不同类型细胞中与受体偶联的细胞内的机制不同所致。

下面我们将重点介绍与细胞表面受体有关的信号调控。

三、三类已知的细胞表面受体

由于细胞表面受体含量极低，还不到整个细胞蛋白的万分之一，因此提纯获得细胞表面受体以研究它们的性质是极困难的。重组 DNA 技术、克隆和表达受体蛋白基因，给细胞受体的研究带来了革命性的转变，大大加速了受体研究的进程。在细胞表面，存在着各种各样的大量受体，目前根据结构和作用方式不同被分成三大类。

大多数的受体属于第一类——离子通道关联受体（ion-channel-linked receptor)。这类受体的蛋白质位于通道的周围，当与信号结合后受体蛋白变构而控制通道的形状，从而改变质膜的离子通透性（有关这类受体及其作用，这里不再详细介绍)。

第二类是 G-蛋白关联受体（G-protein-linked receptor)。它的作用是间接地调节另一种结合在质膜上的靶蛋白的活性。靶蛋白可以是离子通道蛋白或者是一种酶，受体蛋白与靶蛋白之间的相互作用是通过第三种蛋白介导的，这介导的第三者称为“三聚体 GTP 结合调控蛋白”，简称 G-蛋白（trimeric GTP-binding regu-

latory protein，G-protein)，这类受体因此而得名。这类受体与信号结合后通过G-蛋白的介导激活靶蛋白（关于激活的方式途径下面将介绍)。如果靶蛋白是一种离子通道，就改变质膜的通透性；如果靶蛋白是一种酶，就改变一种或几种介导物的浓度，通过这些介导物改变细胞行为或这些产物再作用于细胞内别的蛋白质。G-蛋白关联受体是很重要的一类受体，是一个由同源的七次跨膜的蛋白质组成的大的超级家族。

第三类是酶联受体（enzyme-linked receptor)。这类受体中每个受体含一种跨膜蛋白，它本身就是一种酶或是与酶相联系的，通常是蛋白质激酶。当信号与受体蛋白的膜外部分结合后，它的膜内部分变构而表现出蛋白质激酶活性，或者使其膜内部分相结合的另一种蛋白质具蛋白激酶活性，结果使细胞内的一些蛋白质被磷酸化。

第二节　通过 G-蛋白关联受体进行的信号调控

一、G-蛋白关联受体结构——七次跨膜

G-蛋白关联受体是细胞表面受体中最大的一个家族，在哺乳动物细胞中已发现 100 多种，细胞对大量的多种多样的信号分子作出反应是由这一类受体介导的，这些信号分子包括蛋白质、小肽、氨基酸和脂肪酸的衍生物、多种激素以及神经信号传送器（neurotrans mitter)，带有相同配体的信号分子可以激活这一家族中的不同成员，例如乙酰胆碱（acetylcholine）可激活 5 种以上的 G-蛋白关联受体成员，5-羟色胺（serotonin）最少能激活 15 种。

从 DNA 的序列分析结果推出的氨基酸顺序来看，所有分析过的 G-蛋白关联受体的结构是相类似的，肯定它们在进化上有密切关系。它们都含一条多肽链，肽链中部是亲脂的，由这一亲脂部分反复折叠，进出七次穿过质膜（图 11-1)；近 N 端部分为亲水区域，这部分伸到细胞外面，含有多种信号的结合部位。C 端伸入细胞质中，肽链中位于细胞质中的部分（图中所示，从肽链 N 端起第五和第六跨膜片段之间的环）是主要的三聚 G-蛋白结合部位。

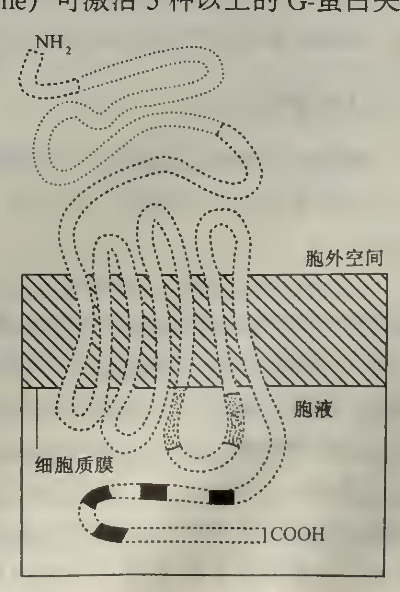


图 11-1　G-蛋白关联受体

二、三聚体 G-蛋白

下面讲的 G-蛋白关联受体作用于靶蛋白的两种主要途径 cAMP 途径和 Ca^{2+} 途径都需要三聚体 G-蛋白介导，因此先介绍三聚体 G-蛋白。它由三种不同的多肽链组成：α、β 和 γ。不同的 G-蛋白 α 亚基是不同的。β 和 γ 亚基形成紧密的复合物，这种复合物通过 γ 亚基的作用而固定在质膜上。当 α 亚基与 GDP 结合时，它就与 βγ 复合物结合形成三聚体，这时的 G-蛋白处在未激活状态。当 G-蛋白关联受体与细胞外的信号结合时，受体构象改变而激活，它与 G-蛋白的 α 亚基结合，α 亚基释放出它原先结合的 GDP 而与 GTP 结合（这时信号与受体分离，受体复原），于是 α 亚基被激活，并与 βγ 复合物分离开。至于 α 亚基再与何种蛋白作用，是激活还是抑制就要看 α 亚基的类型了。如果是 αs 它就与腺苷酸环化酶结合而激活腺苷酸环化酶，结果从 ATP 产生环式 AMP（cAMP）；与此同时 αs 亚基呈现 GTP 水解酶的活性使与之结合的 GTP 变为 GDP，这样带有 GDP 的 αs 亚基与腺苷酸环化酶脱离转而又与 βγ 复合物结合而恢复三聚体状态。

三、G-蛋白关联受体作用的两条主要途径

G-蛋白关联受体接受了细胞外的信号后，通过激活 G-蛋白作用于一种酶，再由酶的作用产生出细胞内的小分子介导物第二信使，通过第二信使作用于靶蛋白。所以可以说，G-蛋白关联受体实质上是把细胞的信号转变成细胞内第二信使，cAMP 和 Ca^{2+} 就是它间接产生的两种第二信使，图 11-2 叙述这两种途径。

1．cAMP 途径

激活的三聚体 G-蛋白的 α 亚基直接作用于腺苷酸环化酶，腺苷酸环化酶是结合在质膜上的一种蛋白。当 α 亚基与它结合时它变构被激活，从 ATP 合成 cAMP。

cAMP 是在 1959 年被最先发现的一种激素作用的细胞内介导因子，它的功能是在所有原核细胞和动物细胞中作为一种细胞内信号介导多种反应及激活基因表达。受激素刺激，在细胞内的水平可在几秒钟内提高 5 倍，又能很快被 cAMP 磷酸二酯酶水解变成 5′AMP，因此它的水平是在不停地升降变化的。许多细胞外信号起调控作用是通过调节 cAMP 的水平，它们一般是通过 Gs G-蛋白受体（α 亚基为 αs）激活 AMP 环化酶而不是激活 cAMP 磷酸二酯酶。最少有 4 种激素可激活脂肪细胞中的 cAMP 环化酶而介导储存的脂肪转变成脂肪酸。同一种信号分子，例如肾上腺素既可增加也可降低细胞内 cAMP 的浓度，这取决于信号分子结合的受体类型，如果结合的受体不是 Gs G-蛋白而是 Gi G-蛋白（抑制型

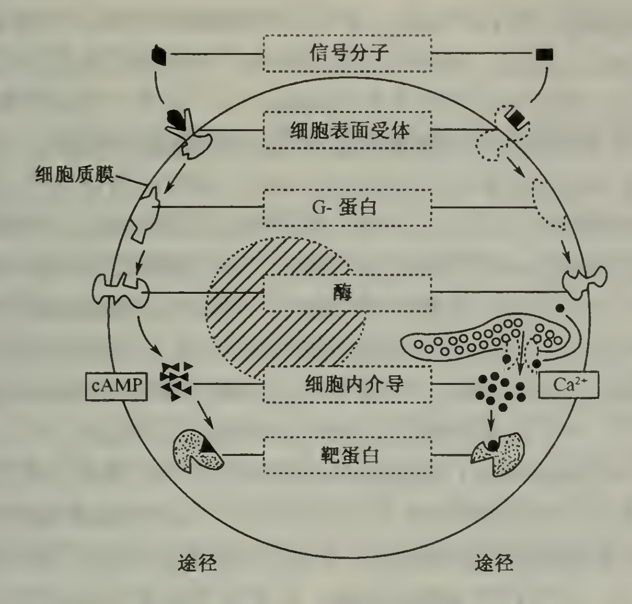


图 11-2　信号通过 G-蛋白关联受体产生小的细胞内介导物的两种途径

G-蛋白，α 亚基为 αi，与 Gs G-蛋白相同），则降低 cAMP 的水平。

在绝大多数动物细胞中，cAMP 激活依赖 cAMP 的蛋白质激酶（A-kinase），这种激酶有选择地催化一些蛋白质磷酸化。催化磷酸化的位置是丝氨酸或苏氨酸，被磷酸化的这两种氨基酸的 N 端与两个以上的碱性氨基酸相连。未激活的 A-kinase 由两个调节亚基和两个催化亚基组成，当 cAMP 与两个调节亚基结合时引起变构激活和释放出催化亚基，使某些蛋白质磷酸化，从而起到调节蛋白质(酶）活性或基因表达（调控蛋白）的作用：例如当 A-激酶使 CREB 蛋白（cAMP response element binding protein）磷酸化后，结合于 DNA 上的 CRE，激活基因转录。

2．Ca^{2+} 途径

受体接受表面信号后激活一种三聚 G-蛋白，称为 Gq。Gq 再激活一种肌苷酸特异性的磷酯酶 phospholipaseC-β（膜结合蛋白），这种酶水解（切割）PIP_2 产生 IP_3 和乙酰甘油。IP_3（肌苷酸三磷酸）是一种可溶性小分子，它离开质膜通过细胞浆迅速扩散，IP_3 结合于内质网上的 IP_3 守门的 Ca^{2+} 释放通道（IP_3-gated Ca^{2+}-release channel），使 Ca^{2+} 从内质网释放出来，进入细胞浆。Ca^{2+} 作为细胞内介导物，以及这一途径中产生的乙酰甘油都能影响细胞代谢和基因表达。目前对 Ca^{2+} 途径的细节尚不是很清楚。

游离的 Ca^{2+} 在细胞质中的浓度是极低的（$\leqslant 10^{-7}$mol/L），但在胞外液体及

在内质网中是较高的（约 10^{-3}mol/L 左右），在细胞浆中的 Ca^{2+} 变化范围在 $1\times10^{-7}\sim5\times10^{-6}$mo/L 之间，维持和调节这种离子浓度梯度显然不是简单的扩散作用，而是靠 Ca^{2+} 离子通道和离子泵的运输。Ca^{2+} 具有广泛的细胞内信使的功能，早在 1947 年就发现在肌肉细胞内注入 Ca^{2+} 可引起肌肉收缩，近年来的研究表明，Ca^{2+} 在介导细胞的多种不同的反应中起了细胞内信使的作用，包括神经传导、分泌和细胞增殖。另外 Ca^{2+} 和乙酰甘油共同参与蛋白质激酶 C（PKC）的激活。哺乳动物细胞中有 8 种以上的 PKC，其中至少有四种被 Ca^{2+} 和乙酰甘油激活、激活的 PKC 可使一些蛋白质磷酸化，例如使 κB 磷酸化而使 NF-κB 与 I-κB分离（I-κB 和 NF-κB 两种蛋白质在细胞质中以二聚体的形式存在），NF-κB 才能进入细胞核内与特异性 DNA 序列结合而激活基因转录。NF-κB 是一种很重要的反式激活因子，不少基因的调控区有 NF-κB 的结合序列。

钙调节蛋白（calmodulin）是普遍存在的细胞内 Ca^{2+} 受体。在所有检查过的动物细胞内存在，每个细胞中约有 10^7 个分子，占全部细胞蛋白的 1%，这种蛋白由一含 150 个氨基酸的多肽链组成，有四个高亲和力的 Ca^{2+} 结合位点，当它结合 Ca^{2+} 后变构。它可以作为酶的亚基，但更重要的是它结合钙后能与多种不同的蛋白质结合而进一步改变构象，改变被结合的蛋白质的活性，受 Ca^{2+}/钙调节蛋白控制的靶蛋白有多种酶类，如膜运输蛋白（如 Ca^{2+}-ATP 酶，把 Ca^{2+} 运出细胞）。其中 CaM-激酶（Ca^{2+}/calmodulin-dependent protein kinase）是重要的一类酶，CaM-激酶 II 与动物的记忆有关。

上述 cAMP 和 Ca^{2+} 两种途径是互相作用，互相影响的。首先，cAMP 和 Ca^{2+} 的水平高低是互相影响的。例如合成和降解 cAMP 的酶受钙调节蛋白控制。反之，A-激酶可以影响 Ca^{2+} 从内质网释放；第二，一些受 Ca^{2+} 或 cAMP 直接控制的酶之间相互影响，如 CaM-激酶被 A-激酶磷酸化；第三，这些酶之间可共用下步反应的靶分子。这样，一些蛋白质就同时受 cAMP 和 Ca^{2+} 的控制。

上面介绍了 G-蛋白关联受体通过三聚体 G-蛋白介导细胞信号反应的两条主要途径。必须记住，有些三聚体 G-蛋白能直接作用于质膜上的离子通道（如 Ca^{2+}、K^+ 的离子通道）而起调节作用，这点不能忽略。

第三节　通过酶关联细胞表面受体进行的信号调控

酶关联受体也是跨膜蛋白，质膜外部分接受信号，细胞质内的部分本身可具酶活性或直接与酶分子结合。已知有五类：①受体 GMP 环化酶；②受体酪氨酸激酶；③酪氨酸激酶结合受体，这种受体与有酪氨酸激酶活性蛋白质直接耦合；④受体酪氨酸磷酸酯酶；⑤受体丝氨酸/苏氨酸激酶。受体 GMP 环化酶被信号激活后直接产生环式 GMP（cGMP）。cGMP 激活依赖于 cGMP 的蛋白质激酶（G-kinase）。它使蛋白质的丝氨酸或苏氨酸磷酸化，以后的级联式反应与 cAMP

相似。许多已知受体属于酪氨酸激酶家族，所以我们在下面着重对这类受体进行介绍。

一、受体酪氨酸激酶是大多数生长因子的受体

在 1982 年，人们发现上皮生长因子（EGF）是一种含 53 个氨基酸的小分子多肽，它能刺激上皮细胞和另外一些类型的细胞增殖。它的受体是一种酪氨酸特异的蛋白质激酶。这种受体是一条跨膜的多肽（一次跨膜），约含 1200 个氨基酸。它们 N 端伸出胞外，占肽链大部分。在 N 端糖基化富含半胱氨酸，肽链的细胞内部分具酪氨酸激酶活性。当 EGF 结合于受体，激活酪氨酸激酶，它就有选择地把 ATP 上的磷酸基团转移到它自己以及别的一些蛋白质的酪氨酸侧链上。许多生长因子与细胞增殖和分化有关，如 FCFS（成纤维细胞生长因子）、HGF（肝细胞生长因子）、NGF（神经生长因子）、M-CSF（巨细胞克隆刺激因子）等。这些因子都是通过受体酪氨酸激酶的作用使信号转移到靶蛋白。

二、形成二聚体是酶关联受体被信号激活的普遍机制

很难想像出一条跨膜多肽的一端与信号结合后如何引起在膜的另一侧的肽链构象改变。后来研究证实当 EGF 与受体结合后，可引起受体的另一端形成二聚体。二聚体的形成是由于受体在细胞质中的多个酪氨酸之间交叉磷酸化。PDGF 本身是二聚体，它能把两个相邻的受体联结在一起（每个亚基结合一个受体）。因此，使受体形成二聚体是信号激活酶关联受体的普遍机制。受体自身磷酸化的酪氨酸结构区域，是下一步被磷酸化的靶蛋白的高亲和力结合区，一旦与受体的这一区域结合，靶蛋白也在酪氨酸残基上被磷酸化而被激活。

对胰岛素和 IGF-I（类胰岛素因子）的信号受体学说则稍有区别，首先表面受体是四聚体，四个亚基通过三个二硫键相连；第二，当受体与胰岛素结合，引起两个亚基在细胞质中的一半变构而不是形成二聚体，使受体的催化功能区域磷酸化，这个功能区域再激活另一种称为胰岛素受体底物-I（IRS）的蛋白质（在 IRS-I 的多个酪氨酸上磷酸化），由 IRS-I 再激活细胞内的信号蛋白。

三、受体酪氨酸激酶上的磷酸化的酪氨酸被具有 SH2 结构的蛋白质识别和结合

已知和激活的受体酪氨酸激酶结合的蛋白质多种多样，例如：GTP 酶激活蛋白（GAP）、磷酯酶（C-γPLC-7γ）、Src 类非受体蛋白质酪氨酸激酶、磷酯酰肌醇 3-激酶（PI_3-kinase）等。对于这些蛋白质的功能了解得不多，但它们中许

多与细胞增殖和癌变有关。尽管它们具有不同的结构与功能，它们都在结构上有两个高度保守的非催化区域，称为 SH2 和 SH3，这样称呼是因为它们最先在 Src 蛋白的同源性区域 2 和 3 发现（src homology region 2 and 3），SH2 区域的功能是与磷酸化的酪氨酸结合，SH3 的功能不太清楚，人们认为 SH3 区域与别的级联反应下游的蛋白质（没有 SH2 结构）结合有关，即 SH3 起蛋白质“接头”作用，使自身没有 SH2 结构的蛋白质通过“接头”作用而间接与受体酪氨酸激酶联系而激活。例如 Sem-5 蛋白，在绝大多数的动物细胞中都有与它同源的蛋白。它们具有 SH2 和 SH3 结构，R 蛋白就是通过 Sem-5 蛋白的作用激活，R 是把信号控制从受体酪蛋白激酶向核内传递的重要蛋白，它与基因表达、细胞增殖和分化有重要关系。在 30％的癌细胞中发现 R 基因变异。在受体酪氨酸激酶信号途径中，许多参与级联式反应的蛋白与恶性肿瘤有关，除上述的 R 外，还有 SCR、RAF、FOS、JUN 等细胞及病毒的癌基因。因此，研究致癌基因有助于鉴定受体酪氨酸酶信号途径中牵涉的各种成分。

第四节　小分子信号调控

一、NO 和 CO 能直接与细胞内的酶结合

虽然信号分子大多数为细胞外亲水的信号分子，它们通过细胞表面受体的介导而起作用，然而尚有些疏水的小得足以穿过质膜的分子如 NO 和 CO、维生素 D 及固醇类激素（甾类化合物，如肾上腺皮质激素，性激素等）、甲状腺激素等也属于信号分子。NO 和 CO 是近年来才认识的气体信号分子。

NO 由精氨酸脱氨产生，由于它是小分子，很易穿过膜在附近细胞中扩散，所以 NO 可在产生它的细胞及附近细胞中起作用，它的半衰期只有 5～10s，只能在局部起作用。NO 能直接与它的靶蛋白（酶）结合而激活酶。人们用硝酸甘油治疗心绞痛已有百年历史，然而到现在才认识硝酸甘油的作用机理，是由于它在转变过程中产生 NO。NO 使血管的平滑肌细胞松弛，增加血液流入心脏而使病情缓解。巨噬细胞和嗜中性细胞也产生 NO 以帮助它们消灭入侵的微生物。许多类型的神经细胞使用 NO 对邻近细胞发出信号。

NO 与鸟苷酸环化酶的活性位点的离子反应，刺激它施放产生 cGTP，我们知道，cGMP 是很重要的细胞内介导物。NO 影响很快，在几秒钟内即能发生。因为 GMP 磷酸二酯酶降解 cGMP 的速度很快。鸟苷酸环化酶也必须很快从 GTP 产生 cGMP。

CO 刺激鸟苷酸环化酶的方式与 NO 相同。

二、维生素 D 和甾类激素等直接和基因转录的调控蛋白结合

NO 和 CO 是直接与酶结合而刺激酶的活性。与上述小分子气体不同，性激素、甲状腺激素、维生素 D 等疏水的小信号分子，尽管它们的结构、功能各不相同，但它们起作用的机制是类似的。它们都穿过质膜而直接与调控某种基因转录的蛋白质结合而使这些蛋白质激活。这些调控基因转录的蛋白的结构通常具有激活转录区域、DNA 结合区域、信号（如激素）结合区域。在无活性情况下，它们与抑制蛋白结合形成复合物，这些抑制蛋白封闭了它们的 DNA 结合部位，当激素与它结合后变构与抑制蛋白分离，暴露出 DNA 结合区域，于是成为激活状态，能与特定的 DNA 序列结合而激活转录（图 11-3）。

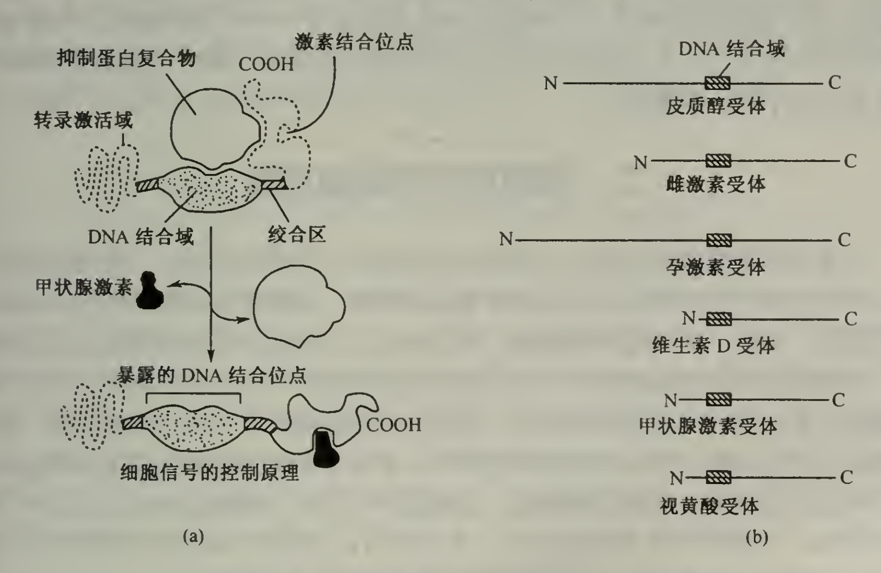


图 11-3　细胞内受体超级家族

（a）细胞内受体蛋白作用模型；（b）各受体的短 DNA 结合域。

这些激活基因转录（偶尔有抑制基因转录的）的蛋白质有的存在于核中，有的存在于细胞质中，当被激活后才进入核中与 DNA 结合。它们结合的 DNA 序列紧挨着被其所激活的基因。这些小分子信号引起的反应在许多情况下可分为一级反应和二级反应。一级反应就是由信号结合而激活的基因转录调控蛋白直接激活的某一基因产物所产生，二级反应是这一基因产物又激活别的基因转录。

这些水不溶性的小信号分子比水溶性信号分子的寿命要长得多，可以介导更长时间的反应。例如大多数水溶性激素在数分钟内即被分解去除，有的信号只能维持几秒钟，但甾类激素可在血液中维持几小时，甲状腺素可维持几天。

第五节　细胞对信号的反应

一、细胞信号逻辑：信号网络

在多细胞动物中，每个细胞要接触各种信号和信号组合，信号调节它们的代谢；信号决定它们改变或继续维持分化状态；信号决定它们是否分裂；信号命令它们生存还是死亡。不同的信号和不同的途径，可互相作用、互相联系。信号在级联式反应中可以同时涉及一种细胞内的中介分子或靶分子。细胞如何能处理这样十分复杂的情况而做到十分协调、准确地做出最佳的反应？显然应该有像计算机网络那样的逻辑系统，以协调处理各种复杂的情况，这就是细胞综合处理各种信号的逻辑-细胞信号网络，以保证多细胞动物成为一个协调的整体。对于这个信号网络如何工作，人们还不了解，可能将来借助生物计算机模拟系统的研究，逐步揭示其中的一些奥秘。

二、细胞对信号的适应性

当各种信号刺激强度变化（即信号浓度变化）范围比较宽时，细胞和机体对待这些信号浓度改变的反应引起的变化是相同的。这就要求细胞对信号适应或降低敏感性。因此，当细胞长期接触一种信号时，它对这种信号刺激的反应就会降低，这种适应性使细胞能够针对信号浓度在很宽范围内的改变而调整它对信号的敏感性。适应性的一般原理很简单：通过负反馈抑制完成。由于负反馈有一段时间推迟，所以如果一种信号刺激突然增加。在起初的短时间能引起较强的反应。适应性可以逐渐减少表面受体数量，这需要好几小时（慢速适应）；也可以快速灭活受体，只要几分钟（快速适应）。另外也可以改变信号途径中级联式反应涉及的一些蛋白（中速适应），所以细胞对信号的适应有快慢之分。

1. 慢适应是细胞表面受体逐渐减少

当一种蛋白质激素或生长因子与细胞表面受体结合后，通常是被受体介导的细胞通过内吞作用进入细胞并被送进内体，大多数受体在酸性的内体环境中卸下结合的信号蛋白。受体本身再回到质膜上重新使用。信号蛋白质或生长因子由溶酶体进行降解。一部分受体在内体中不能释放出信号蛋白而随着信号蛋白一起被溶酶体降解，所以当细胞连续接触高浓度的信号，表面受体就会逐渐降低。这种机制称为“受体下降调节”。

2. 快速适应：受体磷酸化或甲基化

细胞对信号的快速适应常常是与信号诱导的受体磷酸化或甲基化有关。研究

得比较清楚的是肾上腺素受体。当受体与肾上腺素结合，通过 G 蛋白 Gs 激活 AMP 环化酶。当细胞接触高浓度的肾上腺素时，通过两条途径在几分钟内就降低敏感性而适应高浓度的肾上腺素。第一条途径是 cAMP 浓度升高，激活 A-激酶使 β2 受体的丝氨酸磷酸化而干扰受体激活 Gs 的能力；第二条使受体磷酸化的途径是被肾上腺素结合而激活的 β2 受体成为另外一些蛋白质激酶的底物。在受体的 C 末端尾巴多个丝氨酸和苏氨酸磷酸化，这磷酸化的受体尾巴可被一种称为 β 抑制物（β arrestin）的抑制蛋白结合，封闭了受体激活 Gs 的能力。

在研究细菌运动的趋化性中，发现化学引诱物受体的一些酸性氨基酸侧链甲基化，—COOH 上的—H 被—CH_3 代替，这是由甲基转移酶催化的，有的多至 8 个氨基酸甲基化。如引诱物去除，受体甲基化也会解除。受体甲基化也是一种负反馈控制的细胞快速适应性。

上述介绍的细胞对信号适应过程是通过负反馈影响受体的数量和活性。另外的一些适应是通过由信号途径中下游的级联式反应的反应介导物的改变而达到的。例如改变三聚体 G-蛋白就是这种适应途径之一，典型例子是毒瘾。

第十二章　癌分子生物学

第一节　癌发生的分子基础——DNA 序列改变

如果说一个发生了变异的异常细胞经过漫长的道路最终发展成为肿瘤，这个细胞必须能把它的异常性质遗传给它的后代，即这种变异必须是可遗传的。要了解一种癌，首要的问题是看它的可遗传的变异是由于遗传学上的改变即细胞 DNA 序列的改变，还是由于表面遗传的改变，即只改变基因表达的模式但不改变 DNA 序列。这种表面遗传的改变反映细胞的记忆，是我们熟悉的在发育和分化过程中的一种正常现象，它表明分化状态的稳定性。虽然表面遗传的改变也能引起肿瘤，例如畸胎瘤，但我们并不把它作为引起肿瘤的重要原因。因为绝大多数的癌症都是由于 DNA 序列的改变引起的。

致癌因子有三类：①化学致癌剂；②物理射线，如放射性同位素产生的射线、紫外线、X 射线等；③肿瘤病毒。这三类致癌因子都能引起 DNA 序列改变，或引起碱基替换或 DNA 断裂、易位、重组等，结果使细胞发生变异或转化。

在机体中癌的发育和形成往往是好几种条件发生改变和累积所产生的结果。我们说 DNA 序列改变是癌变的分子基础，并不意味着 DNA 上任何一点改变甚至发生在癌基因上的一点改变就能使机体发生癌。事实上，即使在不接触致癌因子的情况下，细胞分裂循环过程中，DNA 复制、损伤、修复过程出现的异常改变的频率约为 10^{-6}变异/（基因细胞）。因此，在人的一生中每个基因经历的变异的机会约有 10^{-10}。从这一观点上看，发生癌变并不奇怪，奇怪的倒是为什么实际上癌发生会那么少。

DNA 上的什么地方发生变异才容易引起癌？癌的一个特点是细胞恶性增生。细胞增殖是直接或间接受调控的，正常的调控基因可以笼统地分为两类，一类的作用是增加细胞的生长和分裂速率，增加细胞的数量；另一类的作用相反，降低细胞的生长和分裂速率从而减少细胞的数量。如果 DNA 上的变异包括插入、转位、放大、碱基替换、缺失等使原癌基因激活而成为癌基因表达出活性更高的蛋白，或者使其正常的蛋白过量表达时，就可能造成细胞的增殖。另一方面，肿瘤抑制蛋白基因缺失或变异，不能产生有功能的肿瘤抑制蛋白时，也可能造成细胞的恶性增生。如果上述两类变异叠加，就会大大提高癌的发生和发展。

第二节　癌的发生和发展包括多种因素的协同作用

一个变异的细胞与癌细胞之间有很大的差别。单是 DNA 上的一点变异是不足以发展成癌的，除非多次变异的累积。从一个 DNA 发生变异的细胞发展成癌，要经过一些中间类型细胞的阶段，这一发展过程要经历一个较长的时期，这一时期中往往包括多种条件和因素的共同作用的累积，有 DNA 变异的累积，肿瘤促进因子的作用，而且生活环境、生活方式、习惯等也有影响，因此，不同的癌在不同国家和地区、不同人群中发生率会有很大差别。从这一观点上看，80％以上的癌是可以避免的。

日本广岛受到原子弹爆炸后引起的白血病，5 年后才明显上升，8 年后达到高峰；生产具致癌作用药品的化工厂工人，接触致癌物质到肿瘤发生要经过 10 年甚至 20 年；白血病病人在发病好几年前可以在病人的白细胞中检查到染色体易位，到发生癌后可检查到进一步的染色体异常，表明是经过再度变异，才发展为白血病的。从细胞的第一次变异到发展为癌的每个中间阶段的细胞，都必须经受环境的选择过程，在机体中重要的是免疫系统和信号。事实上，细胞的增殖和分化受多种途径的综合控制。这些控制途径受遗传和环境的双重影响，肿瘤引发因子（tumor initiator）是指引起细胞 DNA 改变即改变细胞遗传状况的一些因子；肿瘤促进因子（tumor promoter）本身并非诱变剂，即本身并不引起 DNA 序列的改变，但这类物质能加速肿瘤的发生和发展过程。不同的肿瘤促进因子可选择性地加速某种肿瘤的发生。例如佛波酯类物质（phorbol ester）就是一种肿瘤促进因子，它的促癌作用是由于它能激活磷酯酰肌醇细胞内信号途径中的重要的蛋白质激酶——蛋白质激酶 C。一些肿瘤在某些国家或地区的发病率显著高于其他地区，往往与这一地区环境中的致癌因子和肿瘤促进因子的类型和浓度有着密切的关系。

多种因素和条件对肿瘤发生发展的协同作用可以通过对实验小鼠的致癌试验观察到。最容易观察的是皮肤癌，如果在受伤的小鼠皮肤反复涂抹肿瘤引发因子如 DMBA（二甲基苯蒽）后，接着又用肿瘤促进因子如佛波酯类物质 TPA 处理，则可很快观察到小鼠皮肤癌形成，比单用 DMBA 处理在相同时间内发生皮肤癌的比率要高得多。同样，在用癌基因做转基因鼠的实验结果也证实，单独带有 *C-myc* 癌基因的转基因鼠的恶性肿瘤发生率只有 10％，单带有 *v-ras* 的发生率只有 40％，而同时带有 *C-myc* 和 *v-ras* 两种癌基因的转基因鼠，在相同时间内癌的发生率达 100％。除原癌基因变异或因插入、重排、基因放大提高原癌基因产物属性或表达水平外，肿瘤抑制基因的变异，失去正常的抑制肿瘤的功能也是肿瘤发生的重要原因。

第三节 原癌基因和癌基因

原癌基因存在于细胞中，它们是癌基因的等位基因，原癌基因是细胞正常功能不可缺少的，尤其是在胚胎发育阶段。细胞原癌基因的正常功能不会引起肿瘤，但如果原癌基因由于DNA变异被激活造成过量表达，或者由于编码区的变异产生的蛋白质具有更高活性，这时原癌基因就转变为癌基因。癌基因的性质是具有转化细胞的能力，因此判定一个基因是否属癌基因，往往是通过转化实验检查它是否能转化细胞。

癌基因最先是在反转病毒中发现的，就是劳氏肉瘤病毒的 *ras* 基因，后来发现，反转病毒普遍带有癌基因，例如：*src*、*fas*、*mos*、*fos*、*jun*、*myc* 等，而且在细胞中通常存在着这些癌基因的等位基因。除了RNA病毒外，在一些DNA病毒中也带有一些具转化细胞能力的基因，例如大家熟悉的SV40T抗原。DNA病毒的致癌基因通常是早期基因，它们的功能是激活多种病毒和细胞的基因表达或DNA复制，如多瘤病毒的T抗原、HPV的 *E6* 和 *E7* 基因、AdV的 *E1A* 和 *E2B* 基因，也有的是与病毒潜伏有关的基因，如E-BV的 *BNFL-1* 基因。

目前已鉴定出的癌基因约100多种，按照癌基因产物——癌蛋白的功能，可以分为六类。

①生长因子：这类癌蛋白属于分泌性蛋白，分泌到细胞外，可以作用于自身和其他细胞的细胞表面受体。这类癌蛋白的功能与正常的原癌蛋白的功能有何异同，目前还缺少这方面的研究资料。②生长因子受体：这类癌蛋白属跨膜蛋白。除mas外，它们一般能与细胞外的多肽信号分子（生长因子）结合后被激活而具有酪氨酸激酶活性。mas也是一种酪氨酸蛋白质激酶，它的致癌活性是由于它不需要被信号分子激活而能保持高的蛋白质激酶活性。③GTP结合蛋白：这类癌蛋白与质膜相连，具有信号转导功能，即能把表面受体接受的信号向细胞内传递。④细胞内的蛋白质激酶：重要的有酪氨酸激酶、丝氨酸或苏氨酸激酶两类，这类癌蛋白分布在细胞质中；因此用“细胞内的蛋白质激酶”来和跨膜蛋白中的具有蛋白质激酶活性的癌蛋白相区别。⑤信号传递分子（signaling）：这类蛋白分布在细胞质中，具有SH2/SH3结构，可以其信号受体（或蛋白）接受信号，把信号向级联反应的下一级传递。⑥激活转录的调控蛋白：这类癌蛋白分布在细胞核中，常具有HLH、亮氨酸拉链结构（L-Z）。能通过与DNA的相互作用或与其他转录因子的相互作用直接或间接激活转录。属于这类的癌蛋白有C-myc、C-jun、C-fos、C-rel（NF-κB family）以及其他一些病毒癌基因，癌蛋白的分类和分布见图12-1。

从上面所述可以看出，许多癌基因产物是参与信号途径中的各级联式反应，最重要的是传递蛋白质磷酸化的酪氨酸激酶，最终激活转录，导致细胞增殖。另

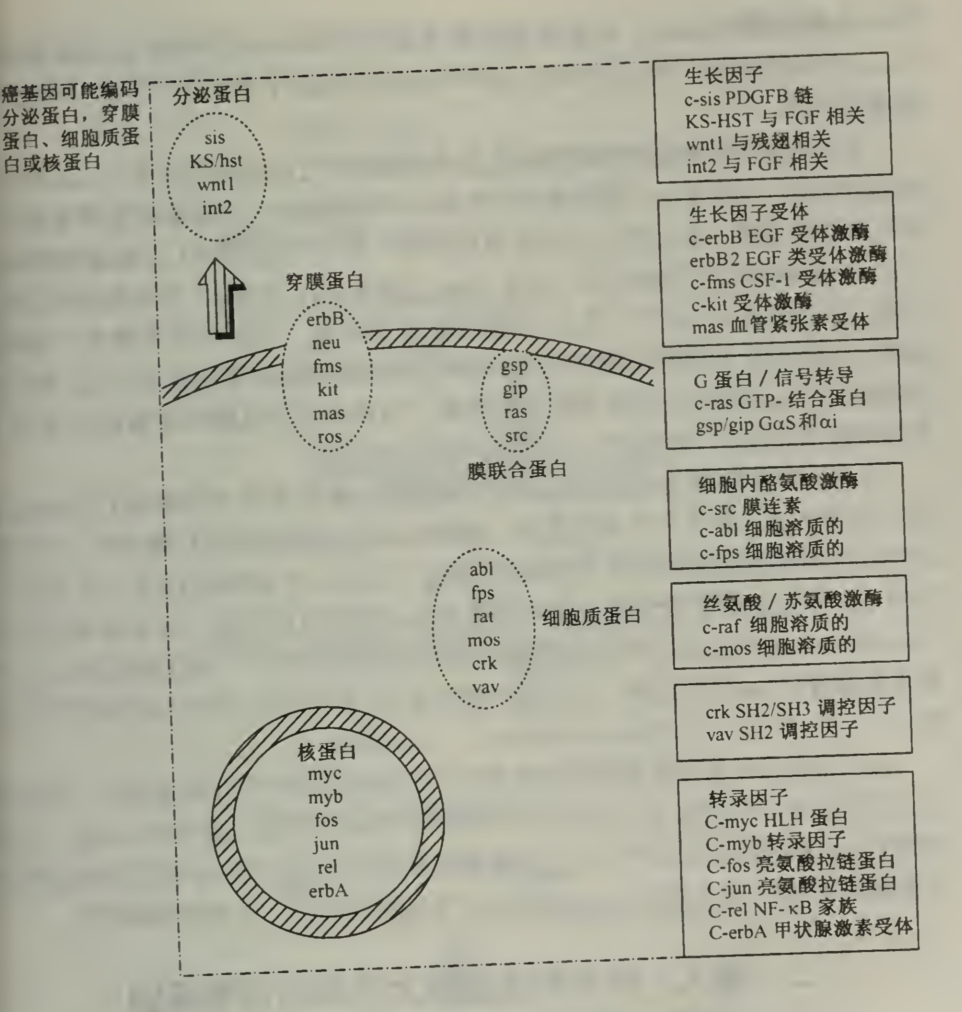


图 12-1　癌基因可能编码分泌蛋白、穿膜蛋白、细胞质蛋白或核蛋白

有一些癌基因产物并不参与信号传递，它们存在于核中直接或间接激活转录，导致细胞增殖。

第四节　原癌基因的激活

一、*ras* 原癌基因可被基因变异所激活

C-ras 家族在人和大鼠中都含三种活动的基因：*N-ras*，*H-ras* 和 *K-ras*，它们分散在基因组中，关系十分密切，它们编码分子质量为 21 000Da 的蛋白质称为 P21ras。Ras 原癌蛋白是一种与质膜相连的 GTP 结合蛋白，在病毒中存在

*C-ras*的等位基因 *v-ras*。在鼠肉瘤病毒（MSV）的 Harvey 株和 Kirsten 株中的 *V-ras*分别与*H-ras*、*K-ras* 同源，每种 V-ras 与 C-ras 十分接近，只有 3 个和 7 个氨基酸不同。

从各种原发性肿瘤或肿瘤细胞系制备的转化性 DNA 中可获得 *C-ras* 的各种癌基因变种。三种 *C-ras* 原癌基因中的每一种都能通过一个碱基的变异而成为具转化性的癌基因。在人的好几种肿瘤中发现，第 12 位或第 61 位氨基酸替换使 *C-ras* 原癌基因转变为癌基因，而在 V-ras（病毒 ras）中第 12 位氨基酸也是变异了的。在正常的 C-ras 中，第 12 位氨基酸是甘氨酸，如果在体外诱变，用除脯氨酸之外的其他 18 种氨基酸中任意一种代替该甘氨酸都能使原癌 *C-ras* 转变成癌基因。C-ras 中的 61 位氨基酸是谷酰氨，如果被除脯氨酸和谷氨酸之外的 18 种氨基酸中的任何一种取代，也能使它转变为癌基因。

上述的氨基酸替换变异使 ras 具有致癌性，显然是转变为癌蛋白后活性改变。ras 蛋白是一种小分子单体蛋白，能结合 GTP 并具有 GTP 酶活性。当它和 GTP 结合时具有活性而作用于它的靶蛋白，之后 GTP 水解成 GDP，与 GDP 结合的 ras 蛋白是没有活性的。变异产生的 ras 癌蛋白可能是一直保持活性状态(constitutive activation)。而保持恒定的激活状态的原因，一种可能性是 GTP 不被水解保持与 ras 蛋白结合，另一种可能性是 ras 蛋白在和 GDP 结合时也具有活性。

除了上述 *ras* 基因在编码区的变异使它从原癌基因转变为癌基因外，也发现在 *ras* 基因的一个内含子（intron）内的变异使 mRNA 加工提高约 10 倍，从而增加了 *ras* 基因的表达水平。在一些肿瘤细胞系中，也发现 *ras* 基因放大造成拷贝数增加 20 倍。当正常的 C-ras 的表达水平显著提高时，也能使细胞转化。

二、插入、转位和基因放大可激活原癌基因

一些原癌基因可以通过改变它们的表达水平而被激活。它们编码的蛋白的氨基酸序列并没有发生改变，已发现的这些原癌基因包括 *C-myc*、*C-abl*、*C-myb*、*C-crbB*、*C-K-ras* 等，其中研究得较为详尽的是 *C-myc*。*C-myc* 提高表达的普遍机制是基因的近邻插入非缺损的反转病毒的 cDNA。

反转病毒在不表达 *V-onc* 基因的情况下仍能转化细胞。首先是在研究 avian 病毒转化的 B 淋巴细胞引起的法氏囊淋巴瘤中发现，类似的情况也在鼠白血病病毒引起的 T 细胞淋巴瘤中发现。在上述两种情况下，反转病毒引起的转化并非是它们编码的蛋白质，而是它们基因组中的 LTR 序列。在许多肿瘤中，病毒基因组 DNA 整合到细胞基因组中，整合的位置在 *C-myc* 基因内或靠近*C-myc* 基因。*C-myc* 基因含三个外显子，第一个外显子代表 mRNA 上长的不翻译的前导序列，另两个外显子编码 C-myc 蛋白质。反转病毒可在相对于 *C-myc* 基因的不

同位置插入，以不同的方式和机理激活 *C-myc*。图 12-2 表示反转病毒 ALV 的 DNA 在 *C-myc* 基因座中不同位置插入和激活 *C-myc* 基因的不同方式。

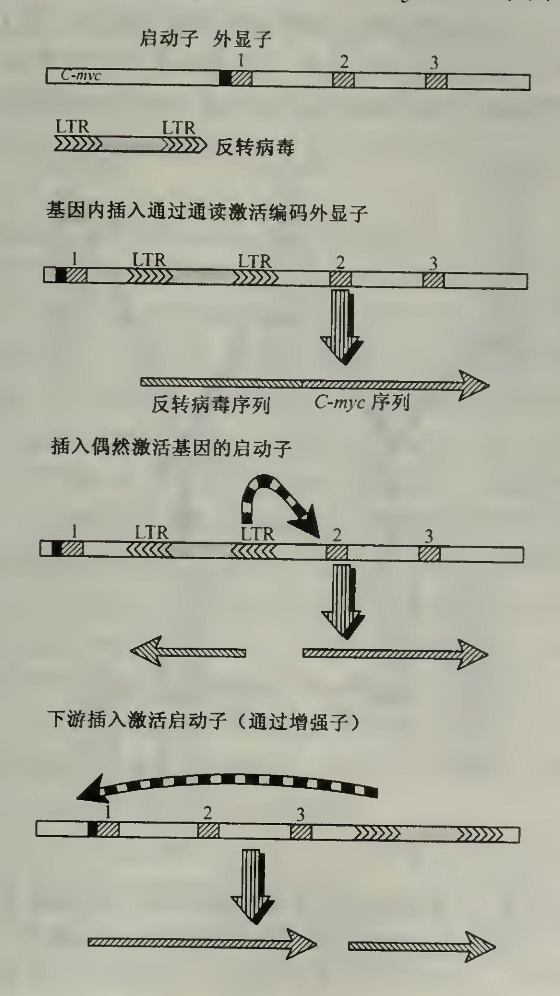


图 12-2　在 *C-myc* 基因座上不同位置插入 ALV 以不同方式激活该基因

当 ALVDNA 插入到 *C-myc* 的第一个内含子，LTR 提供了强大的启动子活性而对两个编码的外显子进行通转录（readthrough），这时与正常的 *C-myc* 表达有两点不同，第一是在强大的 LTR 启动子活性控制下高水平转录，第二是转录缺少了正常的长不翻译前导序列，可能会提高翻译的水平，即在转录和翻译两方面提高了 *C-myc* 的表达。图中另外的插入位置激活 *C-myc* 基因，可能是由于 LTR 的增强子活性或碰巧激活 *C-myc* 的启动子。在上述的情况下，*C-myc* 的编码序列都没有改变，致癌性是由于提高了 *C-myc* 的表达水平。

原癌基因从染色体上的一个位置转移到染色体上的新的位置，是原癌基因激

活的另一种机制。在染色体断裂和基因重排中，如果原癌基因的位置靠近断裂点，重排后被激活的可能性就很大。这种染色体断裂后重接使基因重排而激活基因的现象，起初是在免疫球蛋白基因和某些肿瘤中发现的。图 12-3 是 *C-myc* 基因在染色体上位置改变后导致 B 细胞和 T 细胞的淋巴瘤的情况。第 8 染色体上的 *C-myc* 基因分别靠拢了第 14 染色体上的 *IgH* 和 *TCR* 基因座。

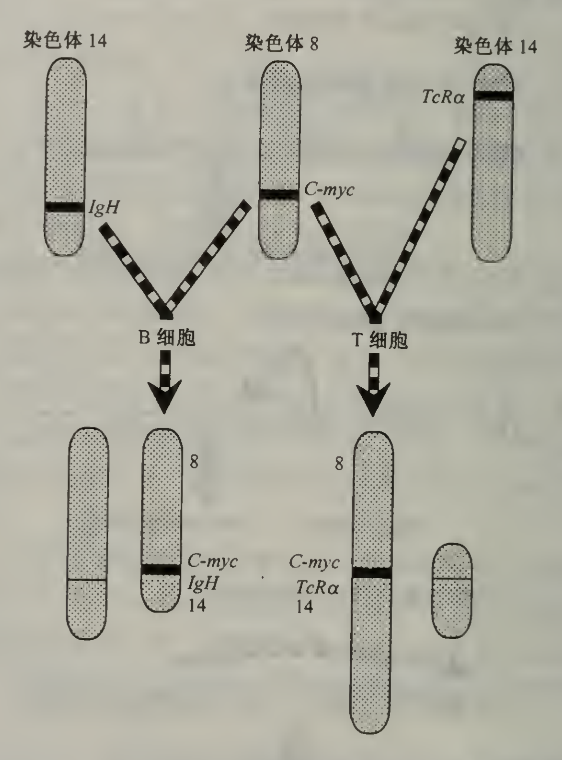


图 12-3　染色体易位是一个相互的事件，两条染色体交换，转位激活人 *C-myc* 原癌基因，包括 B 细胞中的 *Ig* 座位和 T 细胞中的 *TCR* 座位

除第 14 染色体之外，*C-myc* 如与第 2 染色体上的 κ、第 22 染色体上的 λ、第 7 染色体上的 *TCRβ* 基因座靠拢也可能引起肿瘤，这些位置的转移使 *C-myc* 的表达水平提高。例如靠拢 *Ig* 基因座时表达水平被提高 2～10 倍。*C-myc* 基因的致癌性可能暗示，细胞的分化需要 *C-myc* 基因关闭。*C-myc* 基因产物（蛋白质）可能抑制与分化有关的基因表达或激活与细胞恶性增殖有关的基因表达，或两者兼有。由此看来，提高原癌基因 mRNA 的翻译水平如改变 5′的前导序列和 poly A 结构也可以致癌。

许多肿瘤细胞系具有可见的染色体放大区域，在染色体显带（同源染色区）中可以看出，有的放大区不能用显带技术直接看出，通过探针杂交才能检查出基因放大，癌基因放大后造成拷贝数目增加即转录的模板数目增加因而导致mRNA

数目增加。

第五节　肿瘤抑制蛋白

肿瘤的发生有不同机制，原癌基因的激活可引发肿瘤，而肿瘤抑制蛋白基因丢失或变异也是发生肿瘤的重要原因，研究得最详细的是 RB 和 p53 两种肿瘤抑制蛋白。

视网膜细胞瘤（retinoblastoma）是少年儿童的一种病，可以遗传，在视网膜上出现神经胶质瘤。*RB* 基因位于第 13 条染色体上 Q14 带的位置。如果两个 *RB* 拷贝都被灭活就发生这种病。病人常见两个 *RB* 等位基因丢失或一个变异（DNA 缺失一部分），一个丢失。RB 蛋白功能丧失不但可引起视网膜瘤，而且也引起另一些肿瘤，如骨肉瘤、小细胞肺癌。RB 蛋白位于核中，有磷酸化和非磷酸化形式，它对细胞周期有影响。当细胞处在 G0/G1 期，RB 是未被磷酸化的，它可能是在 G1 期末被 Cyclin-CDK 复合物磷酸化的，在细胞有丝分裂期间 RB 去磷酸化。显然，未磷酸化的 RB 蛋白具有防止细胞增殖的功能，未磷酸化的 RB 蛋白能与多种蛋白质相互作用，包括多种癌蛋白和一些转录因子（调控蛋白）。SV40T 抗原、AdV 的 E1A、HPV 的 E7 蛋白在体外都能与 RB 蛋白相互作用形成二聚体。RB 也能与转录因子 E2F、D 型和 E 型胞环素（cyclin）相互作用。因此，RB 抑制肿瘤和影响细胞周期的机制，可能是与癌基因产物（癌蛋白）结合使之失去功能。它与 E2F 的结合可抑制依赖于 E2F 的启动子控制的基因的表达。而对于 RB 蛋白对细胞周期的影响细节，目前尚不大清楚。另有一种肿瘤抑制蛋白——P107，在结构和功能方面与 RB 相似。

p53 是近年来研究得很多的分子之一，它被誉为 1993 年分子。p53 也是一种磷蛋白，最初是在 SV40 转化的细胞中发现的。它在细胞中和 SV40 的 T 抗原偶合成二聚体。在许多转化的细胞和肿瘤细胞系中 p53 蛋白的量大大增加，而且初期的实验把 *p53* 基因的克隆引入细胞中能使细胞永生，导致初期错误地把 *p53* 划归为癌基因，后发现这些 p53 都是变异了的。这些变异的 p53 已经丧失正常的 p53 的功能。p53 没有组织特异性，在许多类型的人癌中可发现 p53 的变异积累，即使在同一种癌，例如乳腺癌中，对不同的乳腺癌病人癌组织中 p53 的 PCR 测序发现，存在不同的 p53 变异，变异集中在 p53 的中部。*p53* 也有两个等位基因，如果其中一个发生变异丧失了功能，就可能引发癌。这点和 RB 不同，在肿瘤病人中一般发现 *p53* 的变异，其中的氨基酸替换比较普遍，也有缺失，很少发现 *p53* 基因丢失。

p53 可能以同源四聚体的形式存在，如果野生型（正常的）p53 多肽和变异的 p53 多肽组成四聚体，那么这种四聚体呈现变异型的构象，没有正常功能。p53 能抑制各种癌基因对细胞的转化作用。它在广泛的组织中起防止细胞增殖的

作用。有两种理论解释 p53 对肿瘤的抑制功能，p53 的一种功能是能和 SV40T 抗原结合，它和 T 抗原的结合可干扰 T 抗原对 DNA 复制的促进作用，而变异的 p53 则不能和 T 抗原结合，推测 p53 能和癌基因产物直接结合而抑制了癌蛋白的活性。p53 的另一种功能是直接作用于 DNA，p53 同时是一种 DNA 结合蛋白，它能识别启动子上 106bp 的对称序列，能激活一些基因的转录，而对于另一些基因，它的结合可抑制基因表达，即它对基因的表达具调控作用，它可能控制与细胞周期有关的一些基因的表达。

p53 的许多变异发生在不同生物共有的保守性区域内，保守的氨基酸变异可能涉及 p53 保守的功能，影响 p53 的性质，有些变异增加 p53 的寿命，半衰期从 20min 增至几小时；有的引起构象改变（可被正常的抗体检出）；有的变异使 p53 不进入核中而存在于细胞质中，不能结合 T 抗原等。p53 的功能缺陷使细胞呈现不同的表形，哪一种功能直接与肿瘤抑制有关，是结合别的蛋白质还是结合 DNA，目前还不能确定。

第十三章　发育与分化调控

发育生物学是一门正在发展中的新学科，它的最终目标是揭开从一个细胞如受精卵变成复杂的多细胞、多组织、多器官的生物之谜，这一过程十分复杂、精细和奥妙。在这个过程中，一系列基因表达是通过级联（cascade）反应来实现的。级联反应是指一些基因的表达产物可激活另一些基因的活性，这些被激活的基因的产物，又调控另外一些基因的表达。这样，最初表达的基因可以逐级地激活和调控其他基因的活性，其效应也就逐级放大。特定基因在特定时间和空间上被激活，产生构成生命有机体的不同组织和特异类型的细胞。可以说，在合适的外界条件下，精确的发育和分化过程是由早已存在于基因组中的程序决定的。在这一章中，我们将集中阐述几种研究得比较清楚的与动物和植物器官或组织发育有关的调控机制，以说明生物在发育过程中基因的表达受严格的时间和空间限制。

第一节　果蝇胚轴的形成

果蝇的卵、胚胎、幼虫和成体都具有明确的前后轴和背腹轴，沿前后轴果蝇胚胎和幼虫显示有规律的分节，从前到后分为头节、3 个胸节和 8 个腹节，在幼虫的两末端又特化为前面的源头和后端的尾节。在早期胚胎发育中，沿背腹轴分化为 4 个区域，即背部外胚层、腹侧外胚层、中胚层和羊浆膜。

一、果蝇胚胎前后轴决定的基因调控

果蝇胚胎前后轴的决定取决于 bcd（bicoid）mRNA 在卵细胞前区的定位和 osk（osker）mRNA 在卵细胞后区极小囊中的聚集。这些 mRNA 翻译产生的蛋白质扩散产生的浓度梯度控制着沿前后轴不同位置的卵裂球内合子基因的表达，从而建立起胚胎的前后轴。

bcd 是一种母体效应基因，在卵子发生中 bcd mRNA 在营养细胞中转录，通过非翻译的 3′区锚定在卵前极而直接进入卵内定位于卵子中预定胚胎的前极，bcd mRNA 使得卵前端成为组织中心。已经发现 bcd mRNA 的定位需要 3 个基因：exuperantia（*exu*），swallow（*swa*）及 staufen（*stau*），其中每一个基因都在定位途径的不同阶段起作用。基因 *exu* 和 *swa* 的突变可破坏卵子发生过程中 bcd mRNA 的定位。stau mRNA 从卵的表面转移到前端的细胞质中，并在卵子发

生后期发挥作用。在发育的开始数分钟内，bcd mRNA 翻译成蛋白质，Bcd 蛋白不能自由扩散，从而建立起 Bcd 蛋白梯度，前极 Bcd 蛋白浓度高，并沿卵的纵轴方向延伸至纵轴一半处。*bcd* 突变可引起 Bcd 产物缺陷，如果两个等位基因都有缺陷（一个同源异型框子无效突变），则幼虫无头和胸，顶节（原头区）即被一个反向的尾节所代替。可引起相似形态的是 bicephalic 突变体，这种幼虫两端各有一个头（称为 Janus 头），且 2 个头呈镜像对称排列，这种具 Janus 头的幼虫是因为卵细胞与两端的滋养细胞邻接，接受两端的 bcd mRNA 产生的。另外，如果将纯化的 bcd mRNA 注射到处于早期卵裂的胚胎中，获得 bcd mRNA 的任何部位都将变成头部；如果 bcd mRNA 被注射到胚胎中央，则中间将变成头部，而它的两边则变成胸部结构；如果把 bcd mRNA 注射到野生型果蝇胚胎的尾部，那么将发育出两端各有一个头的双头胚胎。

Bcd 蛋白结合到卵前区的胚胎细胞核上，它含有一个螺旋-转角-螺旋结构域，由于具有这个结构域，Bicoid 蛋白可以结合到其他基因的启动子上，并把这些基因置于自己的控制之下。

一般认为，Bcd 蛋白梯度提供了位置信息，当卵前部 Bcd 蛋白浓度增加时，头和胸部的边缘都向后移，头和胸的相对大小也增加了，大量 Bicoid 蛋白启动头部发育特异基因的表达，而低浓度的 Bcd 蛋白使得与胸部形成有关的基因特异地表达。Bcd 还可启动合子基因的表达，如 hunchback（*hb*）是首批表达的基因之一，开始，hb mRNA 广泛分布于合胞体卵小室中，但由于受抑制蛋白 Nanos 的影响，它在体内的表达空间被限制在卵前部的 2/3 处。

后部组织中心受母源基因 nanos（*nos*）和 oskar（*osk*）的产物的控制，*nos* 和 *osk* 的产物在细胞后端的定位需要 5 个基因的参与，其中，bicaudal D（*bic D*）、cappucino（*capu*）及 spire（*spire*）失活可阻止 osk mRNA 在卵母细胞定位的第一步——即在前端的聚集，而 *stau* 和 mago nashi（*mago*）的突变可特异性地阻止这个过程后期的一个步骤——即转录产物从前端移到后端。

在野生型胚胎中，Nos 蛋白从卵的后部扩散，形成与 Bcd 蛋白梯度方向相反的浓度梯度，与 Bcd 蛋白不同，Nos 不是一个转录因子，不能结合到 DNA 上，而是通过卵后部 hb mRNA－核糖体复合物抑制转录。

如果前端和后端组织中心都失活，果蝇胚胎仍可产生某些前后图式，这说明还存在第三个与前后轴确立有关的中心，即末端组织中心，在这一组织中心中起关键作用的是 torso（*tor*）基因。*tor* 基因编码一种细胞外信号分子受体蛋白，该信号分子或其前体被卵两端的细胞储存在卵细胞膜和卵周隙中。Tor 受体（一种酪氨酸激酶）被配体占据后能促使身体末端结构、顶节和尾节的形成。*tor* 基因存在两类突变体：一类是失活突变体，胚胎缺失原头区和尾节；另一类是显性突变体，胚胎中部区域（头、胸、腹部）的分节缺失，末端结构却增大。

二、果蝇胚胎背腹轴建立的基因调控

果蝇胚胎背腹轴的建立，涉及约 20 个基因，其中 dorsal（*dl*）等基因的突变会导致胚胎背部化，即产生具有背部结构而没有腹部结构的胚胎。而 cactus 等基因的突变则引起胚胎腹部化，产生只有腹部结构的胚胎。卵子发生时储存了 dl mRNA，但直到受精后 90min，它才翻译成背部蛋白，这种蛋白质开始时均匀分布于整个胚胎，但在腹部信号的诱导下，它们仅定位在胚胎腹部的细胞核中，与核基因结合，形成从腹部到背部的浓度梯度，这种浓度梯度在腹侧组织中可激活腹部化基因（如 snail 和 twisted）的表达，同时抑制背部基因如 *dpp* 和 *zen* 的表达，进而指导腹部结构的发育，建立起胚胎的背腹轴。如果背部蛋白进入所有细胞核中，则胚胎所有细胞都具有腹部的表型。

在正常胚胎的囊胚层中缺乏 Dl 蛋白的细胞将释放一种因子到背侧卵黄周隙中。这种因子是 *dpp* 基因编码的 Dpp 因子，属于生长因子 TGFβ 家族。细胞外 Dpp 梯度和细胞内 Dl 梯度把细胞囊胚再分成几个不同区域：沿腹中线是未来中胚层区；两侧是神经外胚层，将形成神经系统；靠背外侧是背部外胚层，将来形成幼虫表皮。

三、参与果蝇胚胎体节形成控制的基因

果蝇胚胎体节的形成是通过分节基因把早期胚胎沿前后轴分为一系列重复的体节原基，分节基因的突变可使胚胎缺失某些体节或体节的某些部分。根据分节基因的突变表型及作用方式可分为三类：缺口（gap）基因、成对控制（pair-rule）基因和体节极性（segment polarity）基因，这三类基因的调控是逐级进行的。在细胞囊胚层形成之前的割裂中，首先由母体效应基因（如 *bcd* 和 *nos*）控制 *gap* 基因的活化，其次 *gap* 基因之间互相调节彼此的转录并共同调节 *pair-rule* 基因的表达，然后 *pair-rule* 基因之间相互作用，把胚体分隔成一系列重复的体节，并且进一步控制体节极性基因的表达。*gap* 基因和 *pair-rule* 基因再共同调控同源异型基因的表达，所以，胚盘末期的每一个体节原基都具有其独特基因表达的组合，从而决定每个体节的特征。图 13-1 概括了果蝇体节发育过程中参与体节分化的基因的级联调控模式。

1. 缺口基因

缺口基因（gap gene），包括基因 hunchback［*hb*，在第 1～3、13 和 14 副体节（parasegment）前部表达］、kruppel（*kr*，在第 4～6 副体节的后部表达）和 knirps（*kni*，在第 7～12 副体节中表达）等基因，它们的表达区域为一些较宽的

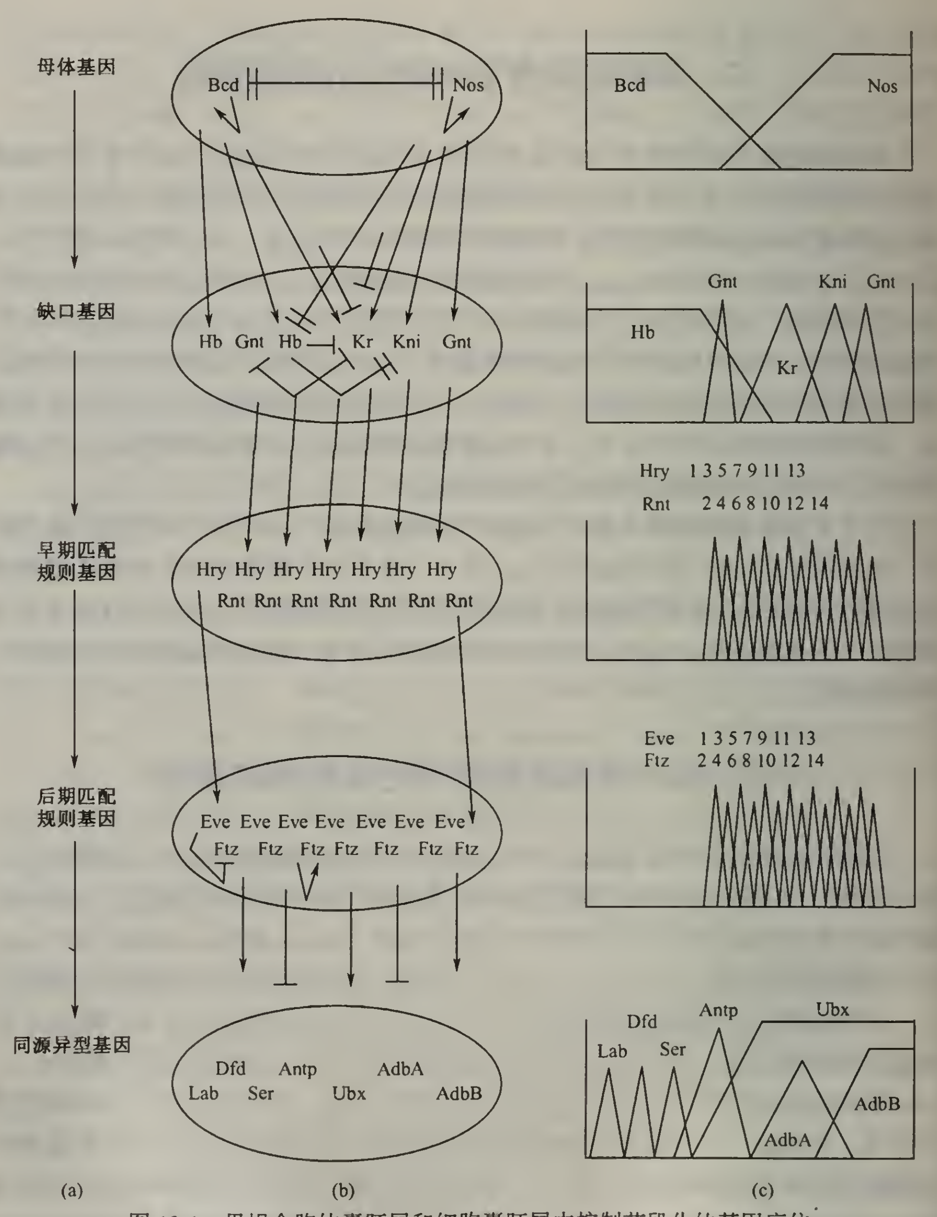


图 13-1 果蝇合胞体囊胚层和细胞囊胚层中控制节段化的基因序位

(a) 基因的序位：母体基因调节缺口基因，缺口基因调节匹配规则基因，按顺序它调节同源异型基因。(b) 箭头表示各种基因产物如何调节基因活性并将基因产物限制在沿前后轴确定的区带。基因活化或蛋白质合成由尖箭头表示（→）；基因的抑制由钝箭头表示（⊣）。双箭头表示强烈活化或抑制(↠或—‖)，在某些情况下低浓度的蛋白质活化基因表达，而高浓度则抑制。这种情况从 Bcd 到 Kr 由两个不同的箭头（→或⊣）表示，蛋白质的符号和名称见表 12.2。(c) 沿着囊胚层长轴的蛋白质梯度显示了基因是如何表达以及基因产物的限制。Bcd 和 Nos 的梯度占据囊胚层各一半，Gnt、Kr 和 Kni 的梯度在囊胚层的后 2/3 被发现。Hry 和 Rnt 交替在囊胚层的后 2/3。这 14 条带大致与形成成虫的节段相对应。Eve 和 Ftz 的 14 条带在囊胚层后 2/3 交替。这些条带确定副节段。

区域，每个区域的宽度约为 3 个体节，表达区之间可有部分重叠，这些基因突变产生的缺陷产物将引起幼虫体内成串的副体节丢失。例如 *kr* 基因突变将引起胚胎对应区域的缺失。缺口基因的表达最初由母体效应基因启动，其表达图式的维持可能依赖于缺口基因间的相互作用。例如 *hb* 基因的表达受 Bcd 蛋白的激活而受 Nos 蛋白的抑制，在它们的相互作用下确定 *hb* 基因在胚胎前部表达。*Kr* 基因同样受到母体效应基因产物的调控，但调节方式不同，Bcd、Nos 蛋白及 Tor 途径均可抑制 *kr* 基因表达。在 bcd 突变的胚胎中，*kr* 基因的表达区域一直到达胚胎前端，当 *bcd*、*nos* 和 *tor* 均失活时，*kr* 基因在全部细胞均表达；而在胚胎中部 Bcd 和 Nos 蛋白的浓度及 Tor 活性均很低，所以 *kr* 基因能够表达。这些缺口基因的表达产物全部是转录调节因子，它们相互调节又具有高度的特异性。一种基因产物可以与其他多种基因的启动子结合，如 Kr 蛋白既能与 *hb* 的启动子结合抑制其表达，又能与 *kni* 基因的启动子结合调节其转录。Tailless 蛋白也识别 *kni* 的启动子并抑制该基因的转录，*hb* 基因产物还可以识别自己的启动子，说明 Hb 蛋白也可以调节自己的表达。

因此，这些基因最初的表达图式通过不同缺口基因之间的相互作用得到进一步的调节和稳定。如 *kr* 基因的表达区域的前端界线处受到 Hb 蛋白的负调控，在后端界线处则受到 Kni 蛋白的负调控；如果 *kr* 基因缺失，*hb* 基因的表达则一直延续到原来 *kr* 基因的表达区域，这种边界性的抑制作用可能是这些基因产物间直接作用的结果。

2. 成对控制基因

成对控制基因（pair-rule gene）：成对控制基因以每隔一个体节，7 条条纹的模式表达，包括许多基因，如 *hairy*、even-skipped（*eve*）、*runt* 和 fushitaraz（*ftz*）等，这些基因的功能是把缺口基因确定的区域进一步分成体节。它们在胚胎期第 13 次核分裂时开始表达，表达图式非常独特，沿前后轴形成一系列斑马纹状的条带，正好把胚胎分成预定体节。如 *eve* 和 *ftz* 基因在胚胎期阶段的表达区域均为 7 条带，只是 *ftz* 在奇数副体节中表达，而 *eve* 在偶数副体节中表达，也就是说，*eve* 和 *ftz* 基因共形成 14 条表达带，这些基因突变将产生仅有奇数或仅有偶数体节的幼虫。成对控制基因由缺口基因活化或抑制，如 *hairy*、*eve* 和 *runt* 的启动子，受缺口基因成对调控，如当 Hb、Kr 和 Kni 蛋白都缺失时，成对控制基因 *hairy* 在整个胚胎中都表达，不呈条纹状区域表达图式，说明在这种情况下，带间区域原来被抑制的 hairy 基因被激活了，如果在成对控制基因表达图式形成的时期向胚胎注射蛋白质合成抑制剂也可获得类似的结果。

3. 体节极性基因

体节极性基因（segment polarity gene）：体节极性基因在每一体节的特定区域细胞中表达，这些基因产物将体节进一步分为 14 个有极性的副节段，当这类

基因突变时每个体节都缺失一个特定的区域。Engrailed（*en*）和 wingless（*wg*）是最重要的体节极性基因，在原肠作用开始时，*en* 和 *wg* 的表达已经开始，但在原肠作用过程中成对控制基因的表达量减少。*en* 基因在每一副体节最前端的一列细胞中表达，*wg* 基因的表达区域刚好位于 *en* 基因表达带之前，即每一副体节的最后一列细胞，因此，这两个基因表达区域的界线正好确立副体节的界线。*en* 和 *wg* 表达的维持则依赖于对方的基因活性，*wg* 基因只在 *en* 基因表达区相邻的一列细胞中维持表达，而 *en* 基因在 *wg* 基因表达区之后的 1～2 列细胞中表达，*wg* 基因编码一种分泌性蛋白，可能直接介导对 *en* 基因表达的维持。体节极性基因 *porc*、*dsh*、*zw3* 和 *arm* 等也都参与这一过程。*wg* 基因编码一种转录因子，*wg* 基因表达的维持由 hedge-hog（*hh*）基因介导，*hh* 基因编码一种分泌性蛋白，表达区域与 *en* 基因的一致。当 *hh* 基因突变时，*wg* 基因表达缺失，与 *en* 基因失活的情况相似。*ptc*、*fu* 和 *slp* 基因在 Hh 蛋白调节 *wg* 基因表达的过程中起作用。其中 *ptc* 编码一种跨膜受体蛋白，它是 Hh 蛋白功能的负调节因子，当 *ptc* 失活时，*wg* 基因表达范围扩大。*en* 和 *wg* 基因的表达受到成对控制基因 *eve* 和 *ftz* 等的激活。*eve* 和 *ftz* 等基因都编码转录调节因子，这是调节 *en* 和 *wg* 的基础。*en* 基因在 Ftz 和 Eve 的浓度达到一定阈值之上时才能被激活，而 *wg* 基因的活化可低于这一阈值。但如果只有 Eve 和 Ftz 还不能激活 *en* 和 *wg* 基因的表达，后者的表达还需要成对控制基因 paired（*prd*）和 odd paired（*opa*）的参与。*prd* 基因最初的表达图式为 7 条较宽的条带，后者很快变成 14 条窄的条带，正好与 *en* 和 *wg* 基因表达区域一致。虽然 *prd* 基因在每个副体节界线处均表达，但起作用的区域与 *eve* 的表达区域一致。所以在 *prd* 基因表达且 Eve 的浓度较高的区域 *en* 基因被激活；而在 *prd* 基因表达且 Eve 和 Ftz 的浓度均较低的区域 *wg* 基因被激活。*opa* 基因可能与 *prd* 基因具有类似的表达图式，而其功能区域可能与 *prd* 互补。在 *opa* 表达和 Ftz 浓度高的区域激活 *en* 基因的表达，而在 Ftz 浓度低的区域激活 *wg* 基因。

4. 同源异型基因

同源异型基因最终决定身体体节将出现哪一种类型，例如一个指定的体节是成为一个无翅的前胸，还是有翅的中胸或有平衡器的后胸？负责做出决定的基因是同源异型选择基因或同源异型主导基因。在果蝇中大多数同源异型基因位于第三染色体上，主要集中在两个区域。一个区域叫触角足复合体 Anrennapdia (Anpt-C)，包含同源异型基因 labial（*lab*）、antennapedia（*antp*）、sex comb reduced（*scr*）、deformed（*dfd*）和 proboscipedia（*pb*）。*lab* 和 *dfd* 基因特化为头部体节。*scr* 和 *antp* 的作用在于使胸部体节一致，特化有翅的中胸需要基因 *antp*，在触角足基因 *antp* 显性突变体中，该基因在头和胸部表达，因此头的部分转变为胸体节的特征，应该长触角的地方却长出了一对附肢。同源异型基因的

第二个区域是双胸复合物 bithorax（bx-C），在这个复合物中，有三个蛋白质编码基因：ultrabithorax（*ubx*），它对于第三胸节的一致性是必需的。abdominal A（*abdA*）和 abdominal B（*abdB*）负责腹部体节的同一性。当 *ubx* 基因缺失时，第三胸节（特征是具有平衡棒）转变成另一个第二胸节，结果是形成一个拥有 4 个翅膀的果蝇。包含 antp 复合物（antp-C）和 bithorax（双胸）（bx-C）复合物的染色体区经常被看成同源异型复合体（Hom-C）。

antp-C 和 *bx-C* 基因排列在第三对染色体上，它们在染色体上的直线排列顺序大致反映了它们表达上的时间和空间次序。沿果蝇身体从头向后移动，首先看到 antp-C 蛋白，并且这些蛋白质继续出现，在后面的胸部，第一个是 Bx-C 蛋白，在腹部最后一个体节 A8 出现由 Hom 复合体的几个基因编码的蛋白。

第二节　果蝇体细胞的性别分化

20 世纪初，一系列遗传学实验结果确定了“果蝇的性别由 X 染色体和常染色体数的比率（X∶A）决定”这一重要概念，即 X∶A 的比值是个体性别决定的

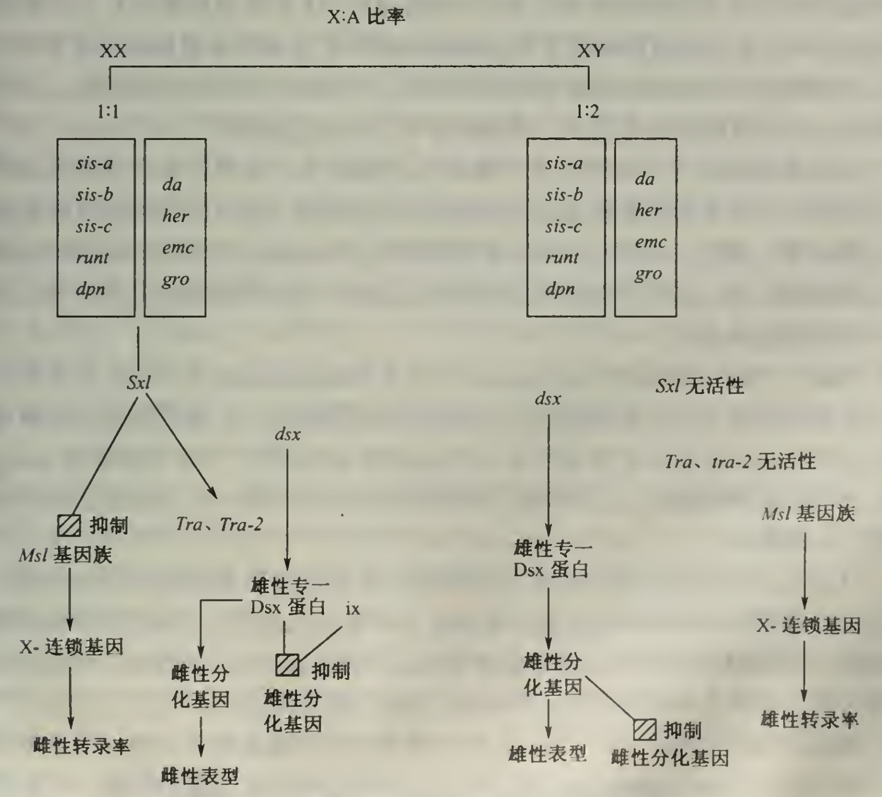


图 13-2　果蝇体细胞性别决定级联调控通路模式

初始信号。

突变分析表明，基因 sex-lethal（*sxl*）、transformer（*tra*）和 transformer-2（*tra-2*）突变可使 XX 个体变成雄性，但这种突变对 XY 雄性个体无作用；基因 intersex（*ix*）突变可使 XX 果蝇成为间性者；基因 doublesex（*dsx*）在两种性别的分化过程中都起重要作用，*dsx* 缺失则可使 XX 和 XY 个体都发育成间性。那么这些基因是如何在果蝇体细胞性别分化过程中发挥作用的呢？除了这些基因，果蝇体细胞的性别分化又涉及哪些基因？图 13-2 概括了果蝇性别决定过程中参与性别分化的基因的级联调控模式。

一、*sxl* 基因——果蝇性别决定的枢纽

果蝇性别决定的第一阶段涉及 X∶A 的比例信号，在体细胞性别分化过程中，X∶A 比例信号是通过性致死基因 sex-lethal（*sxl*）起决定作用的。*sxl* 基因有早期 PE 和晚期 PL 两个启动子，X∶A 信号首先直接作用于 *sxl* 基因的 PE 启动子，在雌性胚胎中启动 *sxl* 的早期转录，*sxl* 基因 PE 启动子的最初激活就发生自合孢体囊胚形成期（受精后的 2h，合子核通过 12～13 次快速核分裂），*sxl* 基因的早期功能是 XX 胚胎启动雌性发育途径和维持两个 X 染色体适量转录水平所必需的，对体细胞及其相应的剂量补偿过程来说，一旦这个遗传开关被确定，就不可逆地进入雌性或雄性发育模式，不再依赖于 X∶A 比值的存在。

X∶A 是细胞自主的信号，依作用的不同被分为“分子”和“分母”记数因素：所谓分子因素就是那些位于 X 染色体上的基因，增加它们的拷贝数会增大 X∶A 的比率，激活 *sxl* 基因，导向雌性化发育，人为减少它们的拷贝数则会降低 X∶A 的比值，使 *sxl* 无法激活而引起雄性化发育；分母因素指的是具有相反效应的一些常染色体基因。

Cline（1998 年）认为 sisterless-a（*sis-a*）和 sisterless-b（*sis-b*）是主要的两个分子基因，是 *sxl* 的正调控因子，改变两种性别中的 *sis* 基因数量，会使雌性缺乏应有的 *sxl* 活性或者使雄性不适当地激活 *sxl*。另一个分子因素是 *runt* 基因，它的记数作用要比 *sis* 基因弱，在胚胎中的分布不均一，只在胚胎的中央区影响 *sxl* 的激活。

Daughterless（*da*）是一个多效应基因，参与多个发育过程，它的母源产物对 *sxl* 的早期表达起着调控作用，在卵发生时若无 *da* 活性，形成的卵在受精后无论其 X 染色体数量如何，均不能激活 *sxl*，缺乏 *da* 并不影响 XY 雄性个体的发育，但对雌性个体是致命的，使 32X 个体因剂量不适当而死亡。因此，受精后，Sis-a、Sis-b、Runt 和 Da 使 *sxl* 仅在雌性胚胎中转录激活。另一个母本效应基因是 extra macrochaetae（*emc*），与 *da* 的作用相反，它是 *sxl* 的一个负调控子，*emc* 突变对雄性发育是不利的。

一个主要的分母因素是 dead-pan（*dpn*）基因，它是一个母本效应基因，一个具有高比例（sis-b∶dpn）的雄性个体会激活 *sxl* 并致死；相反，具有低比例的（sis-b∶dpn）雌性个体将无法激活 *sxl* 并致死，这两种致死作用均与 *sxl* 不适当的活性或非活性状态有关，但与是否带有母本来源的 *dpn* 多余拷贝无关。另一个分母因素是 extra maacrochr、aetae 的基因（*emc*），也是母本效应基因，其编码的蛋白与 *da* 竞争 *sxl* 启动子结合位点，但与 *da* 的作用相反，它是 *sxl* 的一个负调控子，*emc* 的突变对雄性有害。

上述这些基因，都参与基因转录的调节，所谓 X∶A 信号比例实质上就是激活和抑制 *sxl* 转录的调控蛋白的比例，由这一比例决定 *sxl* 的开和关及维持表达水平，从而决定性别的发育。

二、*sxl* 功能的维持

sxl 的早期启动子 PE 仅在胚胎发育极早期的瞬间转录过程中起作用，*sxl* 基因早期的转录和表达模式均为雌性专一，此阶段的调控发生在转录水平。*sxl* 开始转录后不久，晚期启动子 PL 就被激活，*sxl* 进入 RNA 剪接水平的调控。PL 位于 PE 上游 5kb 处，在两种性别中均被激活，并贯穿随后的发育过程，呈组成型表达。对 sxl mRNA 的 cDNA 的分析结果表明，因 mRNA 差异性剪接，雌雄两性的 sxl mRNA 并不相同，在雌性中，Sxl 蛋白可与自身的前体 mRNA 结合，并将其按雌性方式剪接为成熟 mRNA，编码一个含 354 个氨基酸的蛋白质，该蛋白质与雌性特异的 *sxl* 早期转录产物相同或相近，而在雄性个体中，由于缺乏初始的 Sxl 蛋白，新的 *sxl* 初级转录物按雄性方式进行剪接，RNA 的差异剪接使该终止子包含入雄性特异 mRNA 中，因此雄性特异的转录物在 48 位氨基酸后有一终止密码子（UGA），从而产生一个没有功能的雄性 sxl RNA。在雄性个体中，通过剪接产生的雄性 sxl RNA 包含 3 个外显子，其中终止密码子位于中间外显子中，而在雌性个体中，RNA 剪接只产生 2 个外显子，因为雄性特异的中间外显子作为一个大的内含子被剪接掉。

因此，可以说 Sxl 蛋白的表达始终是雌性专一的，只有在雌性个体中，PL 的转录物才被剪接成能产生有功能的 Sxl 蛋白的 mRNA，这一过程由一个包括 Sxl 蛋白（早期或晚期）的正反馈作用在内的自动调节机制所控制。

Sxl 蛋白在结构上包含 2 个含 90 个氨基酸的 RNA 识别域（RRM），这是 RNA 结合蛋白的特点之一。体外研究也显示 *sxl* 编码的 RNA 结合蛋白（即 Sxl 蛋白）和两个靶目标结合：一个结合靶目标是转化子基因 *tra*；另一个结合靶目标是 *sxl* 前 mRNA，通过这种结合作用，Sxl 蛋白可阻止利用雄性特异外显子的剪接位点，导致雌性发育途径。

sxl 基因产物除了自我调控外，一个更重要的作用是调节下游的性别分化基

因。在体细胞中，这些基因依次是 *tra*、*tra-2*、*dsx* 和 *ix*。

此外还有几个基因是雌性个体中维持 *sxl* 基因正常表达所需的，如 female-lethal-2d［*fl*(2)*d*］基因，它是一个母性效应基因，一旦发生突变，会影响 sxl 晚期转录物的雌性化剪接，其他的还有 *liz*［又称 sans fille（*snf*）］和 virilizer（*vir*）基因。雌性个体中，这些基因中的任何一个突变都将使 *sxl* 因不能维持其高水平表达而无法扮演它在雌性发育中的多重角色。

三、性别分化调控途径中的靶基因

位于 *dsx* 下游的是体细胞性别分化的末端基因，它们在两性中的差异表达确立了成虫果蝇性别的二态性。果蝇中很多蛋白质仅在一种性别中存在，如雌性个体中的卵黄蛋白和卵壳蛋白，*dsx* 对这些靶结构基因的调控可能是在转录水平上进行的。研究表明，*dsx* 的雌雄两种转录物在卵黄蛋白基因 127bp 的增强子内有 3 个结合位点，雄性特异产物与这些位点结合，阻止卵黄蛋白的转录，而雌性特异蛋白通过与这些位点结合并激活转录。除了阻碍雌性特异基因产物和器官的产生，雄性 Dsx 蛋白还可能促进雄性性巢的分化。

性别决定基因的温度敏感型突变株有助于研究者了解特定基因对性别决定开关敏感的确切时间。当注入 *tra*2 基因的温度敏感等位基因（temperature-sensitive，ts）时，果蝇性别发育途径从幼虫晚期到成虫期都处于有活性状态。*tra2ts* 基因为温度敏感等位基因，在许可温度（低温）下表达雌性表型，而在非许可温度（高温）下表达雄性表型。在幼虫晚期和化蛹期，将温度从许可温度提升至非许可温度时，XX 幼虫和蛹发育成雄性。当成虫突变体在低温饲养，成虫的脂肪体将生成卵黄蛋白，并进入卵子。而当此成虫突变体在高温下饲养，卵黄蛋白基因的表达将停止。

第三节　光滑爪蟾胚中组织和轴的发育

卵母细胞的动物-植物极形成归因于在两个半球中缺失不同的母体效应基因。精子进入卵子后，通过引起细胞质的重排而打破原有的辐射对称模式，使胚胎形成背腹极性，同时激活一些植物极细胞中的背部化决定因子（dorsolizing determinant），如碱性成纤维细胞生长因子（basic fibroblast growth factor，bFGF，附着于植物半球的外皮，刺激赤道细胞分化成中胚层）和头蛋白（当它与动物极细胞质接触时，其母本 mRNA 表达，产生的头蛋白位于灰新月下面），这些背部化的植物极细胞被称为 Nieuwkoop 中心（Nieuwkoop center）。Nieuwkoop 中心表达 vgl RNA（专一的与植物极外皮相关）。vgl RNA 编码大量的转化生长因子（transforming growth factor，TGF）家族，可能是光滑爪蟾信号蛋白之一。另

外，头蛋白影响位于赤道（即灰新月所处的位置）邻近的细胞，形成 Spemann 组织者（organizer）。Spemann 组织者刺激原肠胚形成并定位于胚孔的背唇。

在囊胚 5～7h 之间，赤道细胞已分化成赤道中胚层和 Spemann 组织者。产生于 Nieuwkoop 中心的头蛋白被认为可诱导 Spemann 组织者中属于转化因子（TGF）家族至少两种蛋白质的合成。蛋白质 Xwnt 和活化素 B（activin B）合成于 Spemann 组织者。

高浓度的 Xwnt 蛋白质促进原肠胚形成中背侧中胚层的发育，由植物极产生的 bFGF 在中胚层组织的发育中尤为重要。bFGF 和活化素 B（由 Spemann 组织者产生）的梯度共同产生了大量的不同中胚层组织，有助于确定腹背轴。在植物极中，由高浓度骨形态发生蛋白（bone morphogenetic protein-4，BMP-4），bFGF 和 BMP-4 刺激腹侧中胚层的发育。BMP-4 是生长因子 TGF-β 家族的成员。生长因子（以合成的顺序编号）和产生生长因子的组织（VV、VD、Mv、SO 等）发育之间的关系简要概括如图 13-3。

活化素 B 可诱导原肠胚形成和建立前后轴与腹背轴，将活化素 B 的 mRNA 注射到动物极或植物极可诱导不完整的第二个胚。

原肠入口处（原始肠道的腔）高浓度的 Xwnt 和活化素 B 在该区域促进尾的形成，可能与原肠入口处高浓度的活化素能够刺激一种尾发育所需蛋白质短尾（brachyury）的产生有关。而低浓度的这些蛋白质扩散到胚更远的区域，刺激称为 Gossecoid（Gcd）的蛋白质产生，该蛋白质所在区域将变成头。

β-catenin 是一种母本效应基因，是 Nieuwkoop 中心的一个主要细胞因子，其编码的蛋白质 β-Catenin 是一种多功能蛋白，既能锚定细胞膜上的钙黏蛋白，又是个核内转录因子。在爪蟾胚胎发育中 β-Catenin 在预定胚胎背部的积累最初开始于受精时卵质的旋转移动过程。在整个早期卵裂阶段 β-Catenin 继续积累，主要位置仍然在胚胎背部（图 13-4），同时在背部细胞的核仁中也能检测到 β-Catenin。β-Catenin 开始存在的区域包括 Nieuwkoop 中心和组织者区，但是到卵裂晚期只存在于 Nieuwkoop 中心的细胞中。

β-Catenin 对于形成背部结构是必需的，如用反义寡核苷酸消耗 β-Catenin 的转录产物，将引起背部结构的缺失。而将外源性 β-Catenin 注射进正常胚胎的腹侧可诱导次生胚轴产生。GSK-3 在胚轴的形成中也起着非常重要的作用，若在胚胎细胞期注射具有活性的 GSK-3，以后胚发育将受到抑制；若在早期胚胎发育阶段用具有极强副作用的蛋白质抑制腹侧细胞 GSK-3 的活性，可引起次生胚轴形成。糖原合成激酶-3（GSK-3）对 *β-Catenin* 有负调控作用，进而对于背侧细胞分化起抑制作用，这是因为开始在整个胚胎中均分布有 β-Catenin，在腹侧细胞中由于 GSK-3 介导的磷酸化作用而特异降解，而在背侧细胞中由于存在 GSK-3 的抑制子 Disheveled（Dsh）蛋白，所以 β-Catenin 不会降解。*Dsh* 是参与背侧发育的另一个基因，Dsh 蛋白开始存在于爪蟾未受精卵植物半球的皮质部，

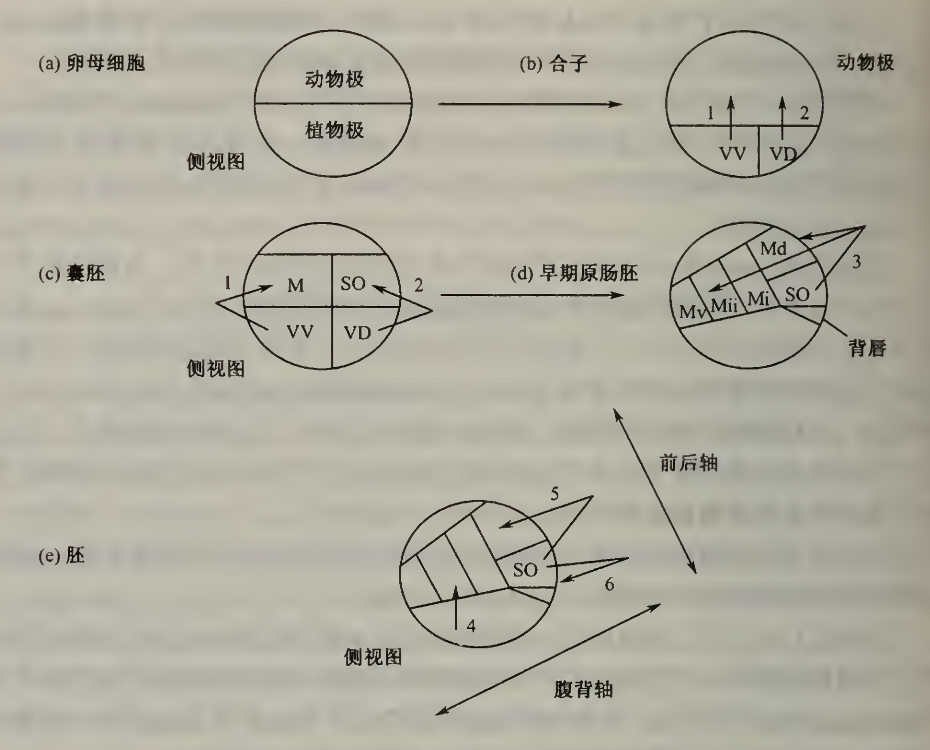


图 13-3　两栖动物中控制组织发育的信号

(a) 两栖动物卵母细胞有由两个半球不同成分确定的动物-植物轴。(b) 卵母细胞受精触发了外皮和与细胞质相关联的胞膜的旋转，产生了灰新月。腹植物极（VV）和腹背区（VD）起始表达不同信号肽。从这些区域引出的箭头表示了生长因子向 overlying 区域的扩散。箭头旁边的数字表示生长因子合成大致的顺序。(c) VV 区的细胞产生碱性成纤维生成因子（bFGF）——第一信号；VD 区（Nieuwkoop 中心）的细胞产生肽段头蛋白（Nog）和转化生长因子（Xwnt）——第二信号。腹侧中胚层（Mv）由 bFGF 诱导，而 Nog 和 Xwnt 诱导 Spemann 组织者（Spemann organizet，SO）的发育和原肠胚形成的起始；(d) 远端胚孔唇的 Spemann 组织者产生肽段活化素 B 和 Xwnt，诱导中胚层的间介型（Md、Mi 和 Mii）——第三信号：(e) Spemann 组织者还建立了胚的腹背轴和前后轴。中胚层和内胚层被另一个称为骨形态发生蛋白（BMP-4）——第四信号的转化生长因子（由 VV 区产生）影响。goosecoid——第五信号——在胚的前区最大表达，短尾——第六信号——主要在生物体的后部产生。

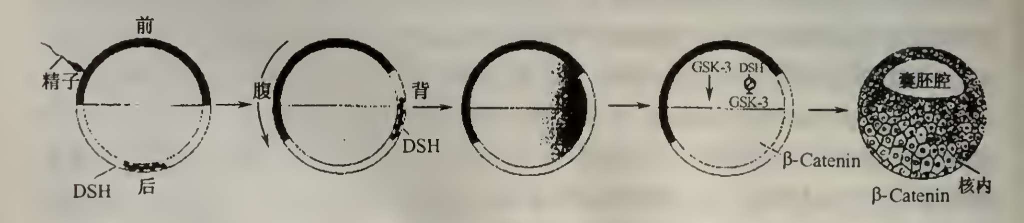


图 13-4　两栖类背-腹轴的形成从精子入卵时开始，
DSH，β-Catenin 和 GSK-3 在背-腹轴形成中具有重要作用

在受精时 Dsh 蛋白沿微管束迁移到胚胎的背侧起作用。

由于 β-Catenin 是转录因子，它可以与其他转录因子结合并赋予后者新的特征。如爪蟾 β-Catenin 与一种普遍存在的转录因子 TCF3 结合，形成的 β-Catenin/TCF3 复合物能够与对胚轴形成具有重要作用的其他几个基因的启动子结合并激活他们的转录。被激活的基因中也包括 siamois（*sms*）基因，该基因于囊胚中期转换后立即在 Nieuwkoop 中心表达。TCF3 是 *sms* 基因转录的抑制因子，但当 β-Catenin/TCF3复合物与 *sms* 基因的启动子结合时可使 *sms* 基因激活。*sms* 基因编码的蛋白 Sms 在组织者特异性基因的表达中起着关键性的作用。Sms 能与 *gcd* 基因的启动子结合并激活该基因表达，其蛋白质产物 Gcd 又为 Spenmann 组织者许多基因的激活所必需。存在于组织者细胞中的 *gcd* 基因是 Nieuwkoop 中心分泌蛋白因子作用的主要靶基因之一。gcd mRNA 表达的区域与组织者的范围有关。如果用氯化锂处理原肠期以前的胚胎可引起组织者中胚层的范围扩大到整个边缘带，*gcd* 基因的表达区域也同样被扩大。与此相反，如果在第一次卵裂前用紫外线照射使胚胎发育缺乏背部和前端的诱导，*gcd* 基因的表达也受到抑制。移植实验进一步表明注射标记 *gcd* 基因的供体细胞能够支配邻近寄主细胞形成背轴。Gcd 是一种 DNA 结合蛋白，具有激活胚孔背唇细胞迁移的属性，能够激活控制脑形成的关键基因 *Xotx2*，使后者在前端中胚层和预定脑外胚层中表达。

除 *sms* 的作用之外，*gcd* 基因的激活也受到定位于植物半球和 Nieuwkoop 中心的 TGF-β 基因家族蛋白产物的协同作用，这些蛋白质由内胚层合成，包括 Vgl，VegT 和 Nodal-相关基因编码蛋白。他们中最重要的是 Nodal-相关蛋白，在囊胚晚期有三种 Nodal-相关蛋白：Xnr1，Xnr2 和 Xnr4，在内胚层中 Nodal-相关基因的表达从背侧到腹侧呈一定的浓度梯度分布。这种浓度梯度是由于定位于背侧的 β-Catenin 与定位于植物极的 Vgl，Vegt 信号相互作用的结果。Nodal-相关蛋白的浓度梯度使中胚层细胞分化，含有大量 Nodal-相关蛋白的中胚层细胞分化成组织者，含有较少 Nodal-相关蛋白的中胚层细胞分化成侧板中胚层，不含 Nodal-相关蛋白或含量极低的中胚层细胞分化成腹侧中胚层。

综上所述，为维持正常的发育在胚胎背部细胞中必须含有 β-Catenin，并使 *sms* 基因表达，Sms 与 TCF-基因家族的蛋白产物协同作用使 *gcd* 基因激活，在 Sms 和 Gcd 等共同作用下才能形成组织者。组织者的形成涉及多种基因的激活，存在于 Nieuwkoop 中心的分泌蛋白激活位于其上方中胚层细胞中一系列转录因子，后者再激活编码组织者分泌产物的一些基因。目前已经发现多种组织者特异性转录因子和分泌性蛋白因子。

第四节　哺乳类体轴建立的基因调控

最近的研究发现，哺乳类和人类受精卵中极体的位置几乎总是与胚胎内细胞

团（ICM）的囊胚腔表面定位在一条线上，且在卵裂过程中这些蛋白被分离到滋胚层中而不是 ICM 细胞中。这意味着，哺乳动物卵子的极性确实影响到胎体的模式形成，且很可能影响胚胎本身。

一、参与背腹轴和前后轴形成的基因

小鼠在原肠作用开始前 12h，位于上胚层前部约 1/3 处的前部肠内胚层（AVE）的中央部分有 VE-1 抗原的出现，还有一些基因的表达，如 *otx-2*、*lim1*、*gcd*、*cerberus-related-1* 及 *hex*。原条形成时，AVE 的最前端，将来发育为心脏的地方，开始表达 *mrg1*。在 AVE 稍后一点，即覆盖着将来形成口部外胚层和前胸的上胚层的 AVE 区域，有 *hesx-1* 的表达。

许多在 AVE 中表达后，并在上胚层衍生物中表达的基因发生突变，将会影响前部发育。例如，在 otx2-/-胚胎中，当卵还呈圆柱形的阶段，上胚层后端的定向似乎是正常的，但肠内胚层却在远端，最后导致胚胎没有中脑和前胸。

当将野生型胚胎干细胞加入 otx2-/-桑葚胚细胞组成嵌合体时，尽管上胚层可以突变型或野生型的形式存在，但不开始前部的发育。当用 otx2-/-胚胎干细胞加野生型桑葚胚细胞组成嵌合体时，将会产生以突变型为主要特征的上胚层，前部的发育可以正确地开始，但不能维持。因此，在胚外组织，可能在它表达的肠内胚层中，前部模式形成正要开始时，otx2 的功能似乎是必需的。

最近的研究发现，小鼠 *cripto* 基因的突变体不能使肠内胚层和上胚层正确定向，AVE 的标志物和前部外胚层的标记物，如 *hesx-1* 基因却定位于远端，应该位于后部的胚外中胚层标记却位于前端。Cripto 蛋白是激活 MAOK 途径的一个信号分子，在斑马鱼中，one-eyed pinhead 基因产物也是属于这一家族，该基因的突变会导致头部发育的缺陷和独眼畸形。在小鼠胚胎中，*cripto* 基因开始在上胚层中表达，但在原肠作用前，其转录本仅限于上胚层的近端边缘区，此处将产生原条。

哺乳动物类似于两栖类胚孔唇的位置是其原条前端的亨氏结。哺乳动物的亨氏结包括许多在蛙组织者中发现的蛋白质，如 Gcd，Nodal，Lim-1 和 Hnf3β 等。*nodal* 基因对原条的起始发育和维持发育是必需的。在原肠开始处，即腹部区域首先看到 *nodal* 基因的表达，此后在原条的最前端发现 Nodal 蛋白，当这个基因被剔除时，发育中的胚胎将缺失原条，且不能进行原肠作用。在原肠后期，*nodal* 基因的表达对胚胎左右轴的形成很重要。

当形成原条的细胞聚集成镰形时，可在这些细胞中发现 gcd mRNA，此后当原条向前移动时，又在亨氏结处发现 gcd mRNA。但当亨氏结退化时，就像在两栖类中的一样，能表达 *gcd* 的细胞保存在头部中胚层和咽部内胚层（脊索前板）中。当头部形成时，在最前端的细胞有 *gcd* 的表达。很明显，它的表达对于激

活与头部形成有关的基因起着关键性的作用。如果鼠胚胎中缺失 *gcd* 基因，可以形成正常的体轴，但不能形成头部。

Gcd、Lim-1 和 Hnf3β 蛋白对于特化前背部中胚层的细胞是必需的，而中间和后面背部胚轴是由 BrAchyury（T）蛋白特化的。脊索的形成和分化要求有 *T* 基因的表达，*T* 基因的突变引起后部胚轴的畸形。最近的研究表明，哺乳动物前后轴的特化也是由包含同源异型盒的基因来完成的。

蝇类和哺乳类动物不仅有相同的同源异型基因的基本类型，而且这些基因在各自染色体上的顺序也非常相似，甚至这些基因的表达形式也遵循相同的规律。小鼠和人类基因组在每个单倍体上包含 4 个拷贝的 *hom-C*（小鼠中是 hox A～D，人类是 hox A～D)。一些哺乳类动物的基因（类似于果蝇的 lab、Proboscipedia 和 Deformed）在前端表达，另一些基因（类似于果蝇 Abdominal-B）在后端表达。由于果蝇 *hom-C* 和哺乳动物 *hox* 基因没有一一对应，所以在这两类动物开始趋异分化时可能发生了独立的基因重复。

二、左右轴形成的基因调控

在小鼠中，inversus viscerum（*iv*）基因突变导致心脏和体位随机分布，相反哺乳动物 inversus of embryonic turning（*inv*）基因的纯合突变使得胚胎体位完全反转。遗传控制左右分化的途径是在不同水平、不同级别等方面经过强加的干扰、精细的平衡才造成最后的体位。早期的干扰可使全部体位随机化或反转，而晚期的干扰仅改变一些器官。人类不动纤毛综合征（immotile cilia syndrome, ICS）导致整个体位的随机分布，表明在发育早期有负责分化左右轴的基因激活，ICS 是由于与微管有关的动力蛋白发生突变而引起的。在原先 iv/iv 品系的鼠中，这种突变在运动区的高保守区引起一个氨基酸的变化，因此这个基因现在称为 left-right dynein（*lrd*)。Lrd 也定位在与基因插入突变 legless 有关的 600kb 的缺失区，legless（*lgl*）是 *iv* 的等位基因。纯合的 *lgl* 突变体没有 *lrd*，且也表现出体位随机分布的特性。哺乳动物左右轴的随机化可能由于细胞间运动的缺失，这与分子的不对称运动是启动左右轴特化的关键这个观点是一致的。

在鼠中，*nodal* 在左边侧板中胚层（LPM）表达，但其行为似乎与两个 TGFβ 家族的新成员 *lefty-1* 和 *lefty-2* 一致。在非常早的体节期，所有 3 个信号分子仅在胚胎左边的细胞群中被诱导表达。*lefty-1* 在左边底板的 5 个细胞中大量表达，而 *nodal* 和 *lefty-2* 在预定腹部体壁的左边侧板中胚层同时表达。而在蛙类、鸡和鼠中，在 LPM 发现 *nodal* 之后不久，就发现 *pitx-2* 的表达，*pitx-2* 的表达区域比 *nodal* 的大，包括左边的 LMP，持续时间长，而且对发展中的中肠管和心脏左侧也有作用。

第五节　拟南芥花的发育调控

高等植物中花的形成经历两个阶段：先是营养顶端和腋生分生组织转变成花分生组织，然后再发育成花。控制营养组织向花分生组织分化的基因主要有 leafy（*lfy*）、apetala-1（*ap-1*）和 cauliflower（*cal*）。Lfy 诱导产生 Ap-1 和 Cal，这后两种蛋白可使营养顶端和腋生分生组织转化为花分生组织。由花分生组织开始的花发育依赖于像动物中同源异型基因那样的基因，包括 *ap-2*、*ap-3* 和 agamous（*ag*），这些植物同源异型基因的突变产生异位（ectopia）的花组织。

一朵典型的花由四个从中心向外的同心环组织构成：心皮（封闭胚珠）、雄蕊（产生花粉）、花瓣［吸引传粉者（pollinator），供传粉者附着］和萼片（保护花在芽中的部分）。同源异型基因 *ag* 在花分生组织的中心表达并影响心皮（中心环）和雄蕊（自中心的第二环）的发育，*ag* 突变可阻止心皮和雄蕊的发育以及配子的产生，这样花组织虽然发育了，但萼片代替了心皮在中心的位置，花瓣代替了雄蕊的位置。

同源异型基因 *ap-2* 在两个外轮生体中表达，并且它可影响花瓣（自中心向外的第 3 轮生体）和萼片（自中心向外的第四轮生体）的发育。*ap-2* 突变可阻碍花瓣和萼片的发育，该花瓣将被另外的雄蕊所取代，而萼片则被心皮所取代。

实验表明上述两个同源异型基因产物 Ag 和 Ap-2 彼此之间互相限制二者在花分生组织中表达的区域，就如同果蝇的缺口基因彼此限制表达一样（图13-5）。如果这两个同源异型基因中的一个丢失，另一个基因的表达就会遍及整个花，并

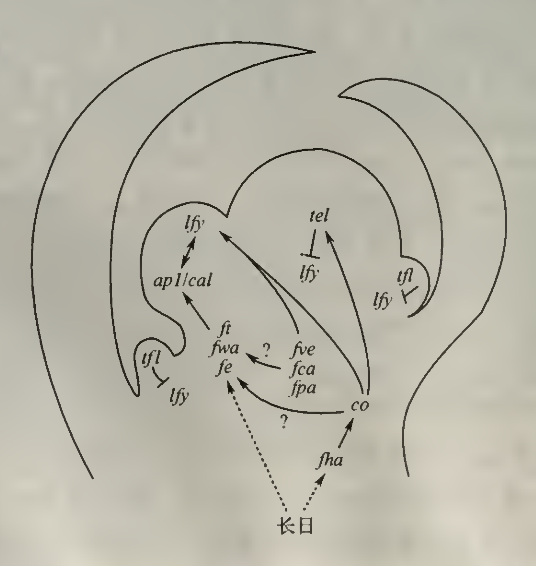


图 13-5　长日对拟南芥成花诱导过程中基因表达的影响

引起环沿着不同的方向发育。遗传学实验和 DNA 结合区的发现表明 Ap-2 是 *ag* 的转录抑制因子，而 Ag 是 *ap-2* 的转录抑制因子。

另一个同源异型基因 *ap-3* 的表达被限制在两个中心环发育的区域。基因 superman的产物使 *ap-3* 保持在中心环之外，可能还有其他基因限制 *ap-3* 在花边缘的表达。*Ap-3* 和 *ap-2* 都是花瓣发育所必须的，*ap-3* 突变导致萼片替代花瓣；雄蕊的发育依赖于 *ap-3* 和 *ag*，*ap-3* 突变导致心皮替代雄蕊。

高等植物成花过程是一个极其复杂的生理过程，受植物内部生长状态和外界环境因子不同方式和程度的影响，那么，控制花发育的基因在花形成过程中是如何发挥作用的呢?

从营养型分生组织转化为生殖型分生组织的发育过程中，基因 *lfy* 和 *ap-1* 起到关键性的调控作用，其他的一些基因如 *cal*、*ap-2*、*ufo* 等可能通过调节 *lfy* 和 *ap-1* 而间接发挥作用。*co*、*emf* 和 *gi* 则为成花计时基因，它们在植物器官中的表达具有数量效应，只有达到某个域值后，才能导致特异性基因的表达。当控制开花时间的基因 *co* 的表达增加到一定的限度时，*lfy* 基因的表达随 *co* 表达的提高而提高；而 *ap-1* 的表达只在有 Ld 存在的条件下提高，这说明 *co* 基因的表达在 *lfy* 和 *ap-1* 之前，且前者可诱导后者的表达，但是 *ap-1* 的表达还需要有蛋白质 Ld 的存在。Co 蛋白可正调控 *ft* 基因，后者可能处于与 *lfy* 基因无关的另一条调控路径上，*ftw* 处在 *ft* 的下游，这两个基因可同时被 Co 激活，并通过不同的调控路径激活花器官特异基因的表达，另外，*tfl* 也能被 Co 诱导。通过序列分析发现，*ft*、*tfl* 与哺乳动物中的 *hncp* 有较高的同源性，Hncp 是一种膜相关蛋白，通过剪切形成游离肽，起到信号传递作用。因此，ft 可能通过蛋白质相互作用而改变其他基因的转录活性，导致不同的基因被转录。在侧生分生组织发育成生殖分生组织的过程中，花原基特异性基因 *lfy*、*ap-1* 和 *cal* 的相互协作有重要作用。*ap-1* 至少通过两条途径被激活，通过原位杂交分析得知，在花分生时期，*ap-1* 的早期转录是由 *lfy* 激活的；而在花器官形成阶段，*ap-1* 的表达在数量上依赖于 *lfy*，说明 *ap-1* 的表达可能是由 *lfy* 间接调控的。而在花器官转变之前或其他营养组织中即使 *lfy* 表达，也不能导致 *ap-1* 的表达，可见 *ap-1* 的转录激活是发生在适当的时间、空间，可能是多蛋白质复合体相互作用的结果。Lfy 是 *ap-*1 的正调节因子，反之，Ap-1 能正调控 *lfy* 基因，因为在转 35S-*ap-1* 的拟南芥中，*lfy* 能在已改变的未成熟的花原基中表达。Tfl 能够通过两条途径抑制花原基特异性基因在营养器官中表达：一条是抑制它们的上调控；另一条是阻碍分生组织对 *lfy* 和 *ap-1* 的响应，因此，在营养分生组织向生殖分生组织转变的过程中，应有一确定时期。*lfy*、*ap-1* 和 *tfl* 基因均表达之后，通过 *lfy* 和 *ap-1* 的正相互作用，以及其他因子的协同作用，使得 *lfy* 和 *ap-1* 表达量增加；而由于 *tfl* 对 *lfy* 和 *ap-1* 具有负相互作用，通过降低 *tfl* 的表达量，使其失去抑制 *lfy* 和 *ap-1* 表达的作用，从而促进花器官特异性基因的表达。*lfy*、*ap-1* 和

tfl 基因在花发育过程中是不可替代的。在花原基分生组织向花器官转变过程中，*lfy* 基因起到关键性的作用，花器官发育的同源异型基因 *ap-1*、*pi* 和 *ap-3* 等都是由 *lfy* 直接或间接调控的。*lfy* 基因在整个花分生组织中都表达，而它所调控的花器官特异性基因却只在特定的花器官中表达，这可能需要一些中间因子同它相互作用，而调控 *lfy* 在不同花器官中的转录，例如在拟南芥中 *ufo* 是 *ap-3* 的上游调控因子，但在 *lfy* 突变体中 *ufo* 没有显著的变化，这些现象表明 *ufo* 不是一个简单的中间因子连接花原基和花器官的发育，它很可能是必需的共调控因子，同 *lfy* 一起作用于花器官基因。研究 *ufo* 同 *lfy* 的相互影响可以作为一个范例，让我们了解像 *lfy* 基因这种在整个分生组织中表达的基因是如何选择性地激活特定区域的基因表达。

第六节　造血系统发育的分子机制

血液细胞可分为两大类，即淋巴细胞（包括 T 细胞、B 细胞和 NK 细胞）和髓细胞（包括粒细胞、单核细胞、红细胞和巨核细胞）。血液干细胞具有向造血系统细胞包括各系髓细胞和淋巴细胞发育的潜能。根据干细胞发育的连续性，血液干细胞只是血液细胞池的一个发育阶段。随着增殖分化，血液干细胞逐渐减弱其自我更新和分化潜能，最终失去干细胞特征，成为没有自我更新分裂增殖能力和功能限定的血细胞。所有血细胞的寿命都很有限，粒细胞仅存活几个小时，红细胞能存活几周，寿命最长的是记忆 T 细胞，也仅存活几年。这样，机体为了维持一定数量的血细胞，就必需持续产生大量血细胞，而在哺乳动物中，所有的血细胞均产生于一种普通的多能干细胞，即造血干细胞。这一部分将主要阐述造血干细胞向各细胞谱系分化及 B 淋巴细胞分化的分子机理。

一、细胞谱系选择的分子基础

一个特定的细胞谱系（如 T 细胞、B 细胞）如何从造血干细胞、多能干祖细胞发育而来是研究造血过程的中心问题。我们知道，不同谱系细胞间的分化或选择过程是基因差异表达的结果。同源盒（*hox*）基因是发育过程中重要的基因，属于转录因子家族，这个家族的成员除在胚胎形成过程中决定细胞命运之外，还与造血细胞形成过程中的细胞谱系特化的控制有关。*hox* 在造血细胞系的差异表达，对造血干细胞的发育、祖细胞的决定或增殖、细胞谱系的特异性分化有着不同的作用。

一些 *hox* 基因如 *hoxB3* 和 *hoxB4* 在含不成熟的造血细胞群中高水平表达。

至今发现至少有 16 个 *hox* 基因（主要是 *hoxA* 与 *hoxB* 簇）在人类骨髓造血干细胞 $CD34^+$ 细胞中表达，在最原始的 $CD34^+$ CD45RA-CD71-造血干细胞

(HSC) 亚群中，*hoxB3* 和 *hoxB4* 高水平表达，在 CD34⁺ 细胞各亚群中，*hoxA10* 基因则等量表达，而在 CD34⁻ 细胞中，*hoxB3*、*hoxB4* 与 *hoxA10* 不表达。

静止期的造血干细胞（HSC）只表达 *hoxB3*，随着分化的进行，位于 3′端的 *hoxB3*、*hoxB4*、*hoxB5* 基因按次序先后表达，而 5′端 *hoxB7*、*hoxB8*、*hoxB9* 在造血干细胞分化的各阶段均未检测到。用原位杂交技术在成年鼠、新生鼠及不同孕期胎鼠的多个部位分析 *hoxB6*、促红细胞生成素（EPO）及其受体基因的表达，发现在造血发生中，三者之间具有时间与空间上的一致性。来自 CFU-E、BFU-E 和 TER⁻119⁺（一种红细胞的特异标记）的红系造血祖细胞中有 *hoxB6* 基因表达，而在 HSC 及非造血系统的干细胞中无 *hoxB6* 基因表达；当后两种细胞分化进入红系系列，*hoxB6* 和红系相关标记又重新表达。因此认为，在红系与粒系的发育分化过程中，特别在 HSC 分化的早期阶段有 *hoxB* 簇基因的表达；在造血细胞分化过程中，*hoxB* 簇基因按 3′向 5′方向逐渐活化。*hoxB3* 是早期造血的主控基因。*hoxB6* 基因与红系造血密切相关，可被视为红系造血细胞的标志。

用针对 hoxB mRNA 的反义硫代磷酸寡聚核苷酸处理造血干细胞（HPC），以阻遏 *hoxB* 基因表达，HPC 形成 BFU-E 与 CFU-GM 的能力显著降低。

用定点阻断（targetedinterruption）造成小鼠 *hoxA9* 基因纯合子型缺失，小鼠外周血白细胞与淋巴细胞总数减少 30%～40%，粒系细胞对粒细胞集落刺激因子（G-CSF）的反应性降低，小鼠胸腺体积显著变小，骨髓中粒系、红系和前 B 祖细胞数目也显著减少；但混合集落、12 天的 CFU-S 和长期培养起始细胞（LTC-IC）则无明显减少。

在造血祖细胞中，*hoxB8* 的表达能特异地影响谱系发育，即抑制粒系发育而促进巨噬细胞系的分化。

hoxB4 能增强 ES 细胞的造血潜能，促进胚胎干细胞（ES 细胞）向造血细胞分化形成红系集落（CFC-E）和粒、红、单核及巨核系混合集落（CFC-GEMM）。*hoxB4* 基因转导的骨髓细胞有很强的扩增原始 HSC 池的能力，使可移植的全能 HSC 数量增加 50 倍，且这种 HSC 的高度扩增不伴随外周血细胞的形态异常，不影响细胞的正常分化。

hoxA10 的过度表达显著地干扰髓系和 B 淋巴细胞系的分化，并导致含巨核细胞和原始细胞的大集落形成。在受致死量照射的鼠体内移植 *hoxA10* 转导的鼠骨髓细胞能使受体鼠重建骨髓造血；在移植后 8～15 周，小鼠髓系干/祖细胞数量明显增加，且有较强的巨核细胞集落形成能力。如果将 *hoxB4*，*hoxA10* 基因转导的骨髓细胞在含生长因子的液体培养基中培养 7 天后，注入受致死量照射的鼠体内，然后测定其 12 天的 CFU-S，与对照组比较，*hoxB4* 和 *hoxA10* 的过度表达可显著增加 CFU-S 数量。研究认为，在造血细胞增殖分化早期，*hoxB4* 基因发挥重要的调节作用；*hoxA10* 的过度表达可导致髓系干/祖细胞的扩增，能改变祖细胞的正常分化。

hox 基因表达对淋巴细胞的分化也有很大影响。G. Saucagvau 等（1997 年）建立了 *hoxB3* 过度表达模型，将 *hoxB3* 基因导入小鼠骨髓细胞并移植进入鼠体内，结果受体鼠胸腺减小，$CD4^+CD8^+$ 细胞绝对数量减少为原来的 1/24，而含 $\gamma\delta TCR^+$ 细胞的 $CD4^-CD8^-$ 胸腺细胞数量增加 3 倍；B 细胞分化紊乱；骨髓与脾脏中粒-巨噬细胞集落形成细胞增加。提示在早期髓系和淋巴系发育中，*hoxB3* 基因的表达能影响其增殖与分化。

可见，*hoxB3* 是早期造血的主控基因，*hoxB4* 能促进胚胎干细胞（ES 细胞）向造血细胞分化，形成红系集落（CFC-E）和粒、红、单核及巨核系混合集落（CFC-GEMM），*hoxB6* 基因是红系造血细胞的标志基因，*hoxB8* 决定造血干祖细胞向巨噬细胞系的分化方向，*hoxA10* 的过度表达可导致含巨核细胞和原始细胞的大集落形成。

骨形态发生蛋白（BMPs）是参与造血干细胞发育的另一个重要的调节因子，它是信号传导生长因子 β（TGF-β）超家族的成员。编码 BMPs 的基因是同源盒基因上游及下游的发育调节网络的一部分。*BMP* 基因在造血细胞系中表达，表达模式与造血干细胞谱系相关，造血细胞能表达如下基因：*BMP-2*、*BMP-4*、*BMP-6*、*BMP-7*、胎盘骨形态发生蛋白（PLAB），生长和分化因子-1（GDF-1），转化因子-β3（TGF-β3）。这些基因的表达具有细胞谱系限制性，*BMP-4* 仅在 T-淋巴细胞系中表达，*BMP-7* 在正常循环的淋巴细胞和肉瘤中的 B 淋巴细胞中都表达，PLAB 在具有分化为单核细胞/巨核细胞能力的细胞系中表达，GDF-1 主要在髓细胞中表达，而 BMP-2、GDF-1 和 PLAB 的表达受细胞分化程度的影响。与正常组织相比，BMP-4 和 PLAB 在白血病细胞系中的表达水平明显升高，表明 BMP 信号传递功能紊乱是引起造血细胞不正常发育的一个因素 。

Adams 及其同事（1992 年）发现，髓细胞 416B 在分化成巨核细胞时表达造血转录因子 GATA-1，同时失去了髓细胞的表型。GATA-1 表达水平也可影响细胞的表型，即 GATA-1 具有剂量效应，如嗜酸性粒细胞表达中等水平的 GATA-1，而成血小板则表达高水平的 GATA-1。NF-M（小鸡的 C/EBPβ 的同聚物）诱导嗜酸性粒细胞的分化，并与 V-myb DNA 结合阻断血小板分化途径。mafB 是 AF-1 超家族的成员，可在骨髓单核细胞中特异表达，并阻止髓细胞向红细胞分化，另外，抑制内源性转铁蛋白受体基因表达也可减弱红细胞的分化。

4-1BB（CD137ILA）是神经生长因子受体（NGFR）肿瘤坏死因子受体（TNFR）家族的成员之一，它介导的协同刺激信号能促进 T 细胞活化、增殖、分化，也能诱导 T 细胞凋亡。4-1BB 既能协同 CD28 对 T 细胞的共刺激作用，也能不依赖于 CD28 而发挥共刺激效应。4-1BB 通过它在胞浆一端的聚集，与肿瘤坏死因子受体相关因子 2（TRAF2）交联，依次活化 ASK1、JNK SAPK 或 p38MAP，启动级联放大效应。活化 B 细胞表达 4-1BB 配体（4-1BBL）。4-1BB 单抗提高 NK 细胞杀伤肿瘤的活性。4-1BB 通过双向信号传导，使单核细胞活

化、增殖，并促进单核细胞的黏附作用。

综上所述，血液细胞的谱系决定是基因差异表达的结果，那么，决定了的细胞谱系又如何发育成执行特定功能的成熟细胞呢？下面以 B 淋巴细胞为例，介绍执行特殊功能的成熟细胞发育的分子机制。

二、B 淋巴细胞发育的分子机制

B 细胞的发育过程分为抗原非依赖阶段和抗原依赖阶段，前一发育阶段指由干细胞→祖细胞→原 B 细胞→前 B 细胞→不成熟 B 细胞，这一发育过程在骨髓中进行；后一阶段指未受抗原刺激的天然（native）B 细胞→激活 B 细胞→浆细胞以及从活化 B 细胞转为静止记忆 B 细胞；在外周淋巴器官中进行。整个发育过程中有许多特征性的基因表达。而这些特定基因的表达是由众多的转录因子调节的。下面主要介绍 B 细胞各发育阶段基因的表达及其调节的转录因子。

1. 抗原非依赖发育阶段基因表达及相关转录因子

在抗原非依赖发育阶段，B 细胞在骨髓中经历祖 B（PrO-B）细胞和前 B（pre-B）细胞，发育成表达表面 IgM 的未成熟 B 细胞。Pro-B 以及定向 B 细胞各阶段都表达 B220（CD45）分子，Pro-B 还表达 CD43、末端脱氧核苷酸转移酶（TdT、RAG-1、RAG-2）、替代性轻链（λ5 和 VPre-B）、CD19、Igα（mb-1）和 Igβ（B29）。Ig 重链基因进行重排，pre-B 细胞 CD43 表达水平下降，后期不表达 TdT，Ig 重链重排成功，并出现 pre-B 细胞受体复合体（由 μ 链、替代轻链、lgα 和 lgβ 组成）。处于分裂周期中的大的 pre-B 细胞 RAG-1 和 RAG-2 在早期的表达水平下降，当进一步分化时，*RAG* 基因又被诱导表达，并出现轻链基因的重排；轻链基因的重排及表面 IgM 的表达标志着 B 细胞抗原非依赖阶段发育的完成。参与 B 细胞这个发育阶段的转录因子有 PU-1/Spi-1、EtS-1、Ikaxos、E2A、EBF、BSAP、NFκB/Rel、Sox-4 等。

Ets 转录因子家族包括 Ets-1、Ets-2、Elf、Pui 和 Spi-B 等多个成员；它们含有 Ets 结构域，通过类似环-螺旋-环方式与 DNA 的嘌呤区结合。这个家族参与 B 细胞早期发育的成员主要有 PU-1 和 Ets-1。PU-1 中的谷氨酸富集区和酸性氨基酸富集区为其调节功能所必需，另外它还有一个介导蛋白质相互作用的 PEST 结构域。在 B 细胞中，Rag-1、Igα、Igβ、Vpre-B、λ5、BTK、bclZ、TdT、CD19 以及 J 链基因的转录调节元件中含有 ETS 结合位点。Vpre-B、λ5 与 μ 链形成前 B 细胞受体（B cell receptor precursor，Pre-BCR），对 B 细胞的发育至关重要。*PU-1* 基因剔除的小鼠缺失 B 细胞，在胚胎期死亡。*PU-1* 的调节作用需要其他辅助因子的参与，只有当它和干扰素调节因子（*IRF*）家族的 *NF-EMS/Pip* 结合后才对 3′κ 链基因增强子（3′κ*E*）有激活作用，*IRF* 家族的另一成员 *ICSBP* 负调

节 PU-1 对 λ 链基因增强子（*AE*）的作用。

Ikaros 属于 *krupple C2H2* 锌指蛋白家族成员；与 *DNA* 上的 *GGGAAT* 结合。*Ikaros* 在细胞早期发育中起关键调节作用，在发育的最早期就开始起作用，参与限定祖细胞向 *B* 细胞方向发育，在它的作用下产生 *B* 前体细胞。它调节的靶基因包括 *RAG-1*、*Igα*、*Vpre-B*、*λ5* 和 *TdT* 等。其中与 *Vere-B*、*λ5* 和 *TdT* 基因结合的位点和 Ets 识别的序列有重叠。敲除 *Ikaros* 基因的小鼠胚胎缺失 B 细胞，而且发育是在 pro-B 之前受阻，其中与 B 细胞定向和分化相关的两种受体型酪氨酸激酶 flk-2 和 c-kit 分别不表达和表达降低；如果小鼠表达具有显性负调节作用的 Ikaros，免疫系统的发育将完全受阻，而且大多数小鼠三周内死亡。另外，最近克隆了 Ikaros 相关分子 Aiolos，它于后期的胚胎胸腺和成年个体的淋巴器官中表达，稍晚于 Ikaros 的表达。最近研究表明，Ikaros 和 Aiolos 可以通过组蛋白脱酰基酶依赖的方式抑制转录。

eza 基因通过不同的剪接方式可产生 E47、E12 和 EZ～E5 几种产物，它们属于碱性螺旋-环-螺旋（base helix loop helix，bHLH）转录因子家族，其中碱性结构域与 DNA 结合，HLH 结构参与蛋白二聚体的形成。EZA 蛋白（E47）与免疫球蛋白基因增强子结合后能激活其转录及基因的重排。*EZA* 基因剔除使 B 细胞发育过程中的 DJH 重排受阻，$B220^+CD43^+$B 前体细胞缺失。在 $EZA^{-/-}$ 的基础上，通过 *E47* 或 *E12* 转基因小鼠表明两者在 B 细胞发育过程中功能有所不同，前者限定向 B 细胞方向发育，后者则为 B 细胞分化所需。EZA 蛋白的功能受 Id 蛋白的负调节。Id 蛋白含有 HLH 结构域，但不能与 DNA 结合，因此与 EZA 蛋白形成的二聚体没有功能。Id 仅表达于 pro-B 细胞中，*Id* 转基因小鼠与 *EZA* 基因剔除小鼠的表型相似。除 Id 外，锌指蛋白 ZEB 也可抑制 EZA 的功能。

EBF 早期 B 细胞因子（early B cell factor，EBF）是在研究 *mbl* 基因的启动子时发现的，其 C 端有 HLH 样结构，介导蛋白二聚体化，N 端含有一类新的能结合锌离子的基序（H-X3-C-X2-C-X5-C），与 DNA 识别相关。EBF 表达于早期 B 细胞发育的整个过程，但在终末分化的浆细胞中不表达。EBF 能与 λE 和 κ 链基因增强子（κE）结合，EBF 和 E47 能协同诱导免疫球蛋白替代轻链 λ5 和 Vpre-B 的表达。EBF 在 pro-B 向 pre-B 发育中起至关重要的作用，*EBF* 基因缺失使 B 细胞的发育阻断在 Pro-B 阶段；这种 $B220^+CD43^+$ 的 Pro-B 细胞表达胚系 μ 链和 IL-7 受体，但当 DJH 没有重排时，Igα、Igβ、替代性轻链、RAG-1 和 RAG-2 的基因也不表达。

BSAPB 细胞特异性激活蛋白（B-cell-specific ativating protein，BSAP）属于配对盒同源域（aired-box homeodomain）转录因子家族成员，在调节 B 细胞早期发育和中脑发育中起十分重要的作用。在免疫细胞中，只有 B 细胞表达 BSAP，在 B 细胞特异性 *CD19*、*λ5*、*Vpre-B* 及 *Blk* 基因的启动子中含有 BSAP 结合位点，但是在 $BSAP^{-/-}$ 小鼠 B 细胞中只有 *CD 19* 表达异常。BSAP 与 ETS 协同调节

Igα 的表达；与 ETS 家族的 NF-αP 竞争 3′CαE 结合位点，起负调节作用；另外，BSAP 还抑制未成熟 B 细胞中 J 链的表达。$BSAP^{-/+}$ 小鼠一般于出生后三周内死亡，这种小鼠中没有 pre-B 细胞、B 细胞和浆细胞，在骨髓中存在 $CD19^-$、$c\text{-}kit^+$、$B220^+$、$CD43^+$、$IL\text{-}7R^+$、DJH 已重排的 Pro-B 细胞，但是胎肝中没有 $B220^+$ Pro-B 细胞。

NF-κB 是一种分布和作用均十分广泛的真核细胞转录因子，因首先发现于 B 细胞 Ig κ 轻链转录调控中而得名。NF-κB 由 Rel 蛋白家族的成员以同源或异源二聚体的形式组成。Rel 家族中至少有 5 个成员与 B 细胞发育相关，包括 RelA (P65)、c-Rel、RelB、NF-κB（P105）和 NF-κB 2（P100)。NF-κB 因能与 κ 链基因内源性增强子（intronic κ enhancer，κEi）结合而被认为是一种 Pre-B 诱导性转录因子；但各种 NF-κB 基因的敲除对 B 细胞发育影响不大。最近发现 NF-κB (RelA 和 c-ReP）对 B 细胞的分化十分必要，一种具有显性负调节作用而又不被降解的 I-κB 在 pre-B 细胞中过量表达，使 κ 转录降低，VκJκ 重排受阻。

Sox-4 是一种 Sry 样的可溶性 HMG 盒蛋白（Sry-like HMG box protein），是调节淋巴细胞的一种转录激活子。Sox-4 于 T 和 pre-B 细胞系和小鼠胸腺中表达。$Soxu^{-/-}$ 小鼠的循环系统发育和 B 细胞的发育受阻，B 细胞停滞在 Pro-B 细胞阶段，胎肝中对 IL-7 刺激发生增殖的 B 前体细胞数量明显减少，增殖能力也严重降低。

上述的转录因子在 B 细胞早期发育的不同时期发挥作用，主要调节着两个过渡点，即限定祖细胞向 B 细胞发育和 Pro-B 向 Pre-B 发育两个关卡（check-point)。目前已知 PUI 和 Ikaxos 在第一个关卡起作用，而 EZA、EBF 和 BSAP 可能在第二个关卡起作用。

2. 抗原依赖发育阶段的基因表达及相关转录因子

进入外周循环的静止 B 细胞首先进入初级滤泡，BCR 与抗原结合后，摄取、加工和提呈胸腺依赖抗原，B 细胞表达 MHC II、B71（CD8)、B72（CD86)、CD40 分子的水平上调。初级滤泡中 B 细胞扩增后形成生发中心，B 细胞与滤泡树突状细胞相互作用，并进行体细胞突变、高亲和力抗体的选择、类型转换和阴性选择，其中 B 细胞所表达的 CD40、CD19、B7-2、LTα、TNF-α、Fas 等分子以及它们相应的配体参与了此过程。当 B 细胞成功地进行了体细胞突变、受体编辑、类型转换、亲和力成熟及负选择后离开生发中心，进一步分化为记忆细胞和浆细胞。参与 B 细胞抗原依赖发育阶段的转录因子有 Oct-1/2、OCA-B/BOB-1/OBF-1、Ets-1、PUI、STATs、NF-κB/Rel、C/EBP、TFE3/MYC、CIITA、Bcl-6 和 BLIMP-1 等。

Octamer 结合蛋白（Oct-1，Oct-2）属于 POU 蛋白家族，它们能与 DNA 中 oct 元件（ATGCAAAT）结合，这种元件存在于所有 Ig 轻链基因和重链 V 基因

的启动子中。Oct-1 广泛分布于各种细胞，而 Oct-2 主要存在于 B 细胞中。Oct-1/Oct-2 的作用受到共激活因子 OCA-B 的调节，OCA-B 特异表达于 B 细胞中，能提高 Oct-1/Oct-2 增强转录的活性及其调节的特异性。OCA-B 与 CIITA 一起激活 *MHCII* 基因的启动子，与 Ets 因子一起激活 *PU-1* 基因的启动子，*CD20* 和 *CD36* 基因增强子中的 oct 元件在成熟 B 细胞的表达中起重要调节作用。*Oct-2* 基因缺失小鼠的 B 前体细胞正常，但 IgM^+ 细胞变少，这种 B 细胞对多克隆有丝分裂原的刺激不反应，但对 T 细胞提供的刺激信号反应正常。*OCA-B* 基因剔除后，BI 亚群细胞和初级滤泡正常，但没有生发中心，成熟的 B 细胞很少，循环的 B 细胞减少，Ig 亚类及其分泌水平异常，小鼠对 TI 和 TD 抗原的反应均降低。由此看来，OCA-B 和 Oct-2 主要作用于 B2 亚群细胞抗原依赖阶段，而对 B 细胞早期的发育影响不大。

Ets 转录因子家族除调节 B 细胞抗原非依赖发育阶段外，在 B 细胞抗原依赖发育阶段也发挥作用。Ets-1 和 PU-1 分别与 μ 链基因增强子（μE）中的 μA 和 μB 结合，它们与 TFE3 协同作用，可能还有 C/EBF 参与。当刺激 BCR 和 CD40 时，Ets-1 与 AP-1 形成复合体，与 3′κE 结合。PU-1 能与 *VH* 和 *Vκ* 基因启动子的嘧啶富含区结合，另外还参与 J 链基因及其本身启动子的调节；Ets-1 激活 B 细胞中 *DRA* 基因的启动子；Elf 是 B 细胞特异性 *TdT* 启动子的一个重要调节因子。由于 Ets 转录因子家族成员功能的冗余性（redundancy）；$Ets\text{-}1^{-/-}$ 小鼠 Ig 水平正常。但有意思的是，Ets-1 的缺失增加了 T 细胞死亡和 B 细胞的分化。Spi-B 在 B 细胞抗原依赖发育阶段中也起十分重要的作用。$Spi\text{-}B^{-/-}$ 小鼠虽然有成熟的 B 和 T 淋巴细胞，但它们的 B 细胞功能和 T 细胞依赖的体液免疫应答严重异常，这种 B 细胞对 BCR 交联的增殖能力很差，而且引起死亡，免疫小鼠的 IgG1、IgG2a 和 IgG2b 水平很低。

STAT 参与细胞信号向胞核内传送与基因转录的活化调节；STAT 与 Jak 组成 Jak/STAT 信号途径，主要参与一些细胞因子、激素和可溶性配体结合相应的受体所介导的信号转导。B 细胞中表达 4 种 Jak 激酶，Jak1、Jak2、Jak3 和 tyk2。其中 Jak3 选择性地表达于血液细胞中。能被 Jak 信号活化的 STATS 有 7 种：STATS1～STATS4、STAT5a、STAT5b 和 STAT6。如 WEHI1231B 细胞 BCR 结合相应配体后，胞内 Jak-1 和 Jak-2 丰度升高，活化的正常人外周血 B 细胞 Jak-3 水平升高。对 BCR 的刺激引起 STAT3、STATS 和 STAT6 的激活。细胞因子引起的 ε 免疫球蛋白类型转换也是通过激活 STAT 实现的。例如，IL4 激活的 *STAT6* 与 *εI* 基因的启动子结合并激活其转录。$Jak\text{-}3^{-/-}$ 小鼠 B 细胞发育在 pre-B 阶段发生障碍；IL-4 不能激活这种 B 细胞的 STAT6 也不能提高 CD23 和 MHCII 的表达，IL-4 与抗 IgM 抗体一起不能引起 B 细胞的增殖。$STAT6^{-/-}$ 小鼠淋巴细胞在IL-4刺激下不能增殖，IL-4 参与的抗体类型转换也被阻断，小鼠经抗 IgD 抗体免疫后不能产生 IgE。$STAT^{-/-}$ 小鼠模型表明，STAT1 参与介导

IFNα/β/γ 传递的信号，为 IFNγ 诱导 IRF-1 和 CIITA 所必需。敲除 *STAT5a* 和 *STAT4* 后未发现 B 细胞有明显的异常。

NF-κB 除在 B 细胞早期发育中发挥作用外，还调节免疫球蛋白类型转换中的基因重排。CD40 与其配体 CD40L 结合后所引起 B 细胞 *NF-κB* 和 *γ1* 基因的转录有赖于 *γ1* 基因启动子中的两个 NF-κB 结合位点与 NF-κB 的结合；在 IL-4 诱导下，NF-κB 与 STAT6 协同激活 *Iε* 基因的启动子；另外，NF-κB 还调节 *Il-6*、*MHCII* 和 *c-myc* 基因的表达，其中 *c-myc* 的表达影响成熟 B 细胞的凋亡。抗 BCR 抗体刺激 WEHI1231 B 细胞后，胞浆中 p50 二聚体水平升高，抑制 *c-myc* 的转录，最后导致凋亡。小鼠缺失 p50 或 p52 后，B 细胞的免疫球蛋白类型转换和成熟 B 细胞的发育受影响。

C/EBP C/EBP（CCAAT/enhancer-binding protein）属于碱性亮氨酸拉链转录因子家族。B 细胞中的 C/EBP 蛋白有 C/EBPa、NF-IL-6、C/EBPβ、Ig/EBP、CRP3、LIP 和 CHOP 等，其中 Ig/EBP、LIP 和 CHOP 是 C/EBPa、NF-IL-6 和 CRP3 活性的抑制因子。C/BBP 结合位点存在于 *μE*、*κE*、*Iε*、*1γ1* 和多种 *VH* 基因的启动子中，B 细胞中的 *CD22*、*Idl* 和 *IL-8* 基因中也有 C/EBP 结合位点。C/EBP 多数情况下是与其他转录因子包括 NF-κB/Rel、STAT6、AP-l、PU-1、AML-1 以及 SP1 协同发挥作用的。基因打靶实验表明，C/EBP 的负调节作用具有重要意义。Ig/EBF$^{-/-}$ 小鼠死于胚胎期，NF-IL6$^{-/-}$ 小鼠的巨噬细胞和 B 细胞功能异常，小鼠脾脏和淋巴结变大，IL-6 过度表达。

TFE3/myc 属于 bHLHZip 蛋白家族，它们均以二聚体形式存在。根据拉链结构域的不同，TFE3/myc 可分为三个亚家族：①TFE3、TFEB、TFEC 和 Mi；②USF I/II；③myc/max。它们之间可形成异源二聚体。*μE* 和一些 *VH* 基因的启动子中存在 TFE3 结合位点，结合于转录起始位点远端的 TFE3 通过 bHLH zip 结构域可与基因转录起始区的序列相结合，介导转录因子远距离的调节作用；*CD23* 也是 TFE3 调节的靶基因。显性负调控基因转染和基因缺失实验表明，TFE3 和 TFEB 调节 *μE* 的作用比 USF 更重要。TFE3$^{-/-}$ 与 RAG-2$^{-/-}$ 嵌合小鼠的 T 细胞和 B 细胞数目正常，但血清 Ig 水平比正常降低 50%～83.3%，这可能是因为接受 T 细胞提供的信号能力受影响。最近发现 *IgH* 基因内部增强子中的 TFE3 结合位点保守性很差；核结合因子（core binding factor；CBF）也与这些位点结合，并能激活相应的基因。CBF 由 α、β 两个亚基组成，α 亚基能与 DNA 结合，而 β 亚基能提高整个蛋白分子对 DNA 的亲和力。CBFα 或 β 亚基基因剔除的小鼠于胚胎期即发生死亡。

CIITA 和 RFX5，MHCII 类分子转录激活子（MHC class II transactivator，CIITA）是从淋巴细胞缺少综合征（bare lymphocyte syndrome，BLS）病人中分离到的一种 *MHC II* 基因的共激活因子，它与 RFXS 协同激活 *MHC II* 的启动子。缺失 *CIITA* 基因小鼠特定组织的 *MHC II* 表达受损。除 MHC II 外，CIITA

还调节与 MHC II 抗原提呈相关的 Ii 链和 *DM* 基因的表达。CIITA 本身不能结合 DNA，它需要其他共激活因子协同作用，包括 OCA-B、TAFII32 等。CIITA 只表达于 MHC II$^+$细胞中，它的表达可由 IFN-γ 诱导。*CIITA* 是通过第 4 个启动子的激活来表达的，启动子 IV 的活化需要转录因子 STATI 和 IRF-1 的参与。STAT1 只有当与 USF1 同时存在时才能结合 GAS 盒，其中 USF1 与 GAS 盒旁边的 B 盒结合。X 盒结合调节因子 5（regulator factors binding to X box 5，RFX5）也是从 BLS 病人中克隆到的。RFX1～RFX5 及 p36 能与 *MHCII* 启动子中的 X 盒结合，RFX1 和 RFX5 分别以同源二聚体或与 p36 形成异源二聚体后与 DNA 结合。RFX5$^{-/-}$小鼠的胸腺髓质、成熟的树突状细胞和激活的 B 细胞仍表达 MHC II 类分子，但胸腺皮质不表达 MHCII，因而不能进行对 CD4$^+$ T 细胞的阳性选择，并引起严重的免疫缺陷；另外，巨噬细胞和静止的 B 细胞也不表达 MHC II 分子。

Bcl-6，C2H2 锌指蛋白家族，是在研究 B 淋巴瘤中的染色体移位和在 Hodgkins 病中的 *Ig* 基因重排时克隆的，它能抑制某些基因的转录。Bcl-6 表达于生发中心，在静止 B 细胞和 T 细胞中表达水平也很高，但随着细胞的活化表达降低。目前尚未发现 Bcl-6 特异性调节的靶基因，与 STAT5 竞争结合 STAT 位点可能是其调节基因转录的一种方式。Bcl-6$^{-/-}$小鼠没有生发中心，也没有亲和力成熟的抗体，并出现因 Th2 细胞因子升高所引起的炎症。

BLIMP-1，B 细胞成熟诱导蛋白（B lymphocyte-induced maturation protein，BLIMP -1）是用扣除杂交方法筛选 IL-5 和 IL-2 诱导分化的 BCLIB 细胞的 mRNA而克隆到的一种转录因子，是小鼠与人 IFN-β 启动子抑制因子 PRD-BFI 的同类蛋白。BLIMP-1 特异表达于成熟 B 细胞和浆细胞中，它的表达足以引起 B 细胞的分化。最近发现 BLIMP-1 能与 *c-myc* 基因表达调节序列中的抑制位点结合，并抑制它在 B 细胞中的表达，由此介导 B 细胞的成熟分化。小鼠缺失 *BLIMP-1* 基因于胚胎期死亡，提示它可能还参与了胚胎的早期发育。

B 细胞发育过程中转录因子及其他调控基因的表达并发挥作用是一个多步骤、调控严谨的过程，随着与发育过程相关转录因子的研究，发育调控的分子机制逐步得以阐明。转录因子调节多种谱系特征性基因的转录，使多能或双能前体细胞在向某个谱系发育时，失去向其他谱系发育的潜能，它们调节的具体靶基因正逐渐被确定下来，这方面的深入研究将有助于全面认识淋巴细胞的发育过程，并为探索整个生命发育过程开辟先河。

第十四章 细胞凋亡

死亡意味着生命的终止，是生物界普遍存在的现象。细胞的死亡，顾名思义，是指细胞生命的结束。发育生物学和细胞生物学研究的焦点是生与死的问题：生即关于细胞的发育、分化和成熟，关于这方面的研究已有一百多年的历史，目前已发展到在分子水平上研究相关基因的表达调控；但长期以来人们一直忽视了对细胞死亡规律的研究。最近这几年，生物学家逐渐认识到细胞死亡特别是程序化细胞死亡即凋亡有其特殊的生物学意义。

第一节 细胞凋亡的概念及意义

一、凋亡的概念

在多细胞生物中，细胞的死亡有两种不同的形式。一种是细胞坏死或意外死亡（necrosis 或 accidental cell death）。它是由于某些外界的因素，比如局部贫血、高热以及物理、化学损伤或生物的侵袭等，造成细胞急速死亡；另一种称为细胞凋亡（apoptosis），这是 1972 年 Kerr 等建议用来描述伴随细胞死亡的一系列形态学上特定的现象，正如秋天枫叶的凋谢，它是指细胞在一定的生理或病理条件下，遵循自身的程序，自己结束其生命的过程，细胞最后脱落离体或裂解成若干凋亡小体（apoptotic body），而被其他细胞吞噬。

细胞凋亡的形态特征与坏死性细胞死亡的形态特征不同。坏死性细胞死亡的形态特征是：膜通透性增加，细胞外形发生不规则变化，内质网扩张，核染色质不规则地位移，进而线粒体及核肿胀，溶酶体被破坏，细胞膜破裂，胞质外溢。这种细胞死亡过程常常引起炎症反应。细胞凋亡是由细胞内在的有规律的变化引起的，它可由某些因素诱导但不是坏死性的改变。这种细胞死亡的特征是细胞首先变圆，随即与相邻细胞脱离，失去微绒毛，胞质浓缩，内质网扩张呈泡状并与细胞膜融合，线粒体无大变化，核染色质密度增高呈半月形，并凝聚在核膜周边，核仁裂解，核固缩，进而细胞膜内陷将细胞自行分割为多个外有膜包裹、内含物不外泄的细胞凋亡小体。由于这种死亡过程不导致溶酶体及细胞膜破裂，没有细胞内含物外泄，故不引起炎症反应和次级损伤，它的结局是被吞噬细胞或邻细胞所识别、吞噬或自然脱落，其生物化学反应主要是细胞核内的 DNA 被内切核酸酶在核小体之间降解，产生若干大小不一的寡核苷酸片段，在琼脂糖凝胶电泳上呈现阶梯状 DNA 区带图谱（DNA ladder），这些区带由 180～200bp 的不同

倍数的寡核苷酸片段组成，这个长度即是核小体间重复单位的大小。有趣的是有些细胞发生凋亡时，染色质 DNA 并不降解，表明 DNA 降解并不是细胞凋亡必不可少的内容（图 14-1）。

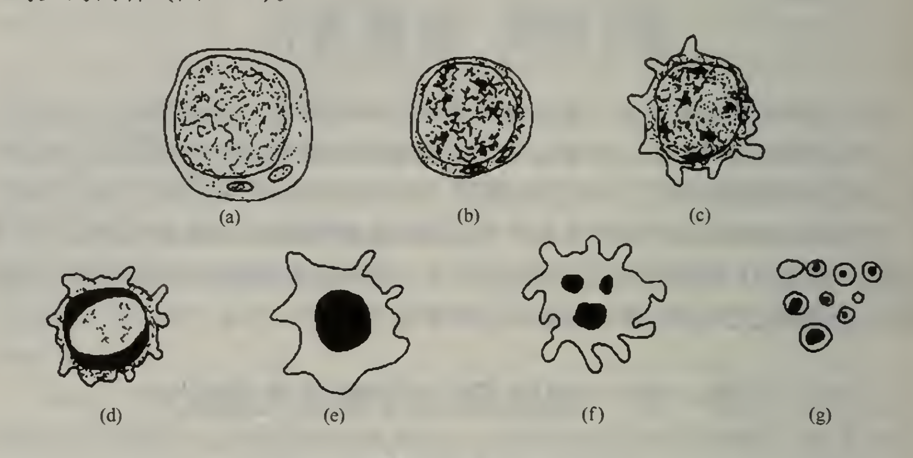


图 14-1　淋巴细胞凋亡过程示意图

(a) 正常淋巴细胞有稀疏的胞质和异染色质。(b) 细胞体积变小，细胞器聚集，染色质凝集为团块。(c) 细胞呈现沸腾运动。(d) 染色质沿核内膜凝集成月牙状。(e) 核凝集为黑孔。(f) 核裂解为小球，细胞凋亡小体发泡。(g) 细胞裂解为凋亡小体。

1980 年，Wgllie 等在糖皮质激素（glucocortocoid，GC）诱导胸腺细胞死亡的实验中，观察到内源性内切核酸酶活化及引起一系列形态学和生物化学方面的变化，提出了程序化细胞死亡（programmed cell death，PCD）的概念。目前，这一概念已为科学家广泛接受。但是不同的科学工作者对 PCD 概念有不同的解释。发育生物学家使用 PCD 来描述在多细胞生物的发育过程中，正常生理诱导的，以时空上不可预料的方式失去细胞的现象，如哺乳动物趾（指）间细胞的丢失，蝌蚪尾巴脱落等。而免疫学家用 PCD 来描述依赖于新的基因表达而引起的任何细胞死亡，与刺激物的性质无关，如 γ 射线，特异抗体和糖皮质激素等引起的细胞死亡。实际上发育生物学家和免疫学家对 PCD 的看法有一点是共同的，即 PCD 需要新的基因表达和蛋白质合成。这一结果已在无脊椎和脊椎动物的细胞培养中使用 RNA 和蛋白质合成或抑制的许多实验中所证实。

大多数程序化细胞死亡，呈现细胞凋亡的形态特征。因此，不少研究人员认为 PCD 和细胞凋亡可以互用，可以等同。但是严格说来，两者指的不是同一现象。前者是功能上概念，而后者是以许多细胞死亡中所见到的固定形式的形态学描述。已经发现，不是所有的 PCD 都表现为细胞凋亡的形态特征，例如，烟草幼鹰蛾变态时，其幼虫体节间的肌细胞由于某些激素的作用而死亡，使幼虫变成蛾。与其他 PCD 类似，这些肌肉细胞的丢失需要新的基因表达，但这些细胞死亡时不产生膜泡、染色质浓缩，DNA 降解等细胞凋亡的共同特征。这一类细胞

死亡称为非凋亡的程序化细胞死亡（non-apoptotic programmed cell death）。

二、细胞凋亡的意义

1. 程序化细胞死亡与个体发育

生物的诞生、发育、成长、存活以及死亡是天然的有规律的程序化过程，其存活至死亡的时间称为寿命。构成生物组织的细胞，其生存和死亡的过程也是天然的规律的程序化过程，也有其寿命。生物整体和细胞虽然都有自然死亡的现象，但两者的生物学意义却有显著的差别。程序化细胞死亡是多细胞动物生命活动过程中不可缺少的组成内容，是动物借以存活的需要，因而贯穿于动物全部寿命周期中。无论是低等动物还是高等动物都是如此。例如某些昆虫从卵到成虫，中间要经过几个蜕变期，每个时期组织结构以及外形都要发生改变，在这些过程中均有赖于新旧细胞的生死交替。死亡的细胞是在完成了它的使命而被淘汰消失的，天然有序。蝌蚪变为青蛙时尾巴自然消失，这种消失的机制是细胞有序凋亡的过程。多种动物的趾（指）的形成，是肢柄端某几部分细胞有序凋亡而被吞噬消除的结果。在哺乳动物胚胎发生、发育和成熟过程中，构成组织的细胞生死交替，程序化细胞死亡是保证个体发育成熟所必需的。PCD 规律一旦失常，个体即不能正常发育，或发生畸形，或不能存活。皮肤分为真皮和表皮，表皮和趾（指）甲是细胞从有生命向无生命的转化，也属于程序化细胞死亡的过程。无此过程，动物即无鳞、无皮、无爪，就失去了保护和防御机制的结构。

人类免疫系统是最有代表性的例子，淋巴系统发育分化成熟过程中，始终伴随着程序化细胞死亡。在 T 淋巴细胞发育过程中，多能干细胞首先发育成原 T（pro-T）淋巴细胞，原 T 淋巴细胞一般没有表面标志。原 T 淋巴细胞进入胸腺后发育成为前 T（pre-T）淋巴细胞，前 T 淋巴细胞主要是 TCR^- $CD4^-$ $CD8^-$的双阴性（DN）细胞。DN 细胞进一步增殖为双阳性（DP）细胞，即 TCR $CD4^+$ $CD8^+$细胞，这是胸腺中最大的细胞群。在这一阶段，T 淋巴细胞抗原受体（TCR）基因开始重排，如果在 *TCR* 基因某一连接点上发生等位基因无意义突变，不能产生或产生不正确的 TCR 分子，细胞即走向程序化死亡。产生了正确的 TCR 分子的 DP 细胞，还必须经过进一步的严格选择机制，使可能识别自身胸腺基质细胞的，即有导致自身免疫病危险的细胞走向程序化死亡，这就是胸腺的阴性选择作用。95%以上的 DP 细胞都是在胸腺皮质区内凋亡，凋亡小体被胸腺上皮细胞吞噬。少数 DP 细胞进一步发育成为单阳性（SP）细胞，即 TCR $CD4^+$ $CD8^-$或 TCR $CD4^-$ $CD8^+$的细胞，这就是成熟的 T 淋巴细胞表型。一些不成熟的 SP 细胞也可以发生程序性死亡。胸腺微环境中细胞和其他因子的不平衡分布可以调节胸腺细胞对程序化死亡的敏感性。在许多情况下，细胞的存活依赖于生长因子的存在，从环境中去除生长因子时细胞则走向程序化死亡。实验证

明，IL-1，IL-2，IL-3，EPO和CSF都有抑制细胞发生程序化死亡的作用。

B淋巴细胞的发育过程与T淋巴细胞相似，前B淋巴细胞发育成为成熟的B淋巴细胞过程中，编码免疫球蛋白的基因片段要经过基因重排，才能在细胞表面产生独特的个体型免疫球蛋白受体。如果发生错误重组，分泌型免疫球蛋白（sIg）就不能正常表达。很难精确计算多少前B淋巴细胞可发育为成熟的B淋巴细胞，但一般来说，免疫球蛋白基因只有按LV/LJ、HV/D和D/HJ的方式连接才能形成正确的密码，产生正确的sIg，在抗原的刺激下，引起克隆消除，从而把可能引起自身反应性的B淋巴细胞克隆清除掉，这就是B淋巴细胞的阴性选择机制。重排的免疫球蛋白转基因对鼠的研究表明，如果这种小鼠表达同种抗原，则机体内几乎没有B淋巴细胞。B淋巴细胞这种消除过程也是通过程序化细胞死亡而实现的。不成熟B淋巴细胞株WEHI-231的细胞表面只有IgM，而没有IgD，当使用抗原免疫蛋白的抗体刺激时，即走向程序化细胞死亡，核小体降解为DNA片段。

成熟的淋巴细胞，也包括其他成熟白细胞的寿命以天计算，死一批，再生一批，互相交替，非常严格有序。若程序化细胞死亡发生障碍，只生不死，就会出现白细胞堆积，发生白血病。该死的不死还会发生自身免疫病。可见这一程序失常可能是血癌和自身免疫病的病因之一。

人的红细胞分化成熟后是无核有功能结构的，存活约120天，然后走向死亡，被吞噬细胞清除。这种有序变化，在程序化细胞死亡过程中的红细胞才有此功能，这是程序化细胞死亡中的一个特殊现象。

上述事例清楚地说明，程序化细胞死亡是多细胞生物保证个体正常发育成熟和维持正常生理过程所必须的。

2. 程序化细胞死亡与肿瘤

在肿瘤的研究中，人们一直在试图阐明肿瘤细胞是如何接受异常的增殖信号而使其得以无限增殖的，至今已取得了不少进展。20世纪80年代以来，人们越来越认识到，很多肿瘤不仅由于其细胞增殖率很高，而且由于其细胞死亡速度很低，即细胞增殖与死亡平衡失调造成的。在肿瘤组织中，通常存在两种细胞死亡的形式，即坏死性细胞死亡和细胞凋亡，后者是由于程序化细胞死亡造成的。存在于肿瘤组织中的凋亡小体和正常组织发育过程中产生的凋亡小体具有相同的超微结构。肿瘤细胞的增殖率与程序化细胞死亡的速率处于一种平衡状态，此平衡状态决定了肿瘤生长的速率。因此，在肿瘤发生过程中如何实现人为地调控程序化细胞死亡的速率是当前肿瘤研究的一个焦点。

使用切除小鼠前列腺降低雄激素的方法诱导程序化细胞死亡，发现雄激素既可促进细胞增殖，又可封闭程序化细胞死亡。去除雄激素后，即使细胞处于G0期也可走向程序化死亡，而且，使用*p53*基因剔除小鼠的研究表明，去除雄激

素诱导的程序化细胞死亡过程与 p53 的功能无关。

有实验表明，程序化细胞死亡参与了肿瘤的起始过程，并对肿瘤的发生起负作用，癌前病灶中程序化细胞死亡率比周围正常组织高 8 倍左右，由于肿瘤细胞分裂周期短，必须通过程序化细胞死亡机制及时地清除老化细胞，肿瘤细胞增殖过程中也可能产生了一些表型近乎正常的细胞，需要通过程序化机制加以清除，从而维持肿瘤细胞群的高增殖率。此外，有些间接的证据表明，正常肝细胞癌变过程中，癌前细胞的程序化细胞死亡更加敏感，容易被清除，这是机体自我保护功能的表现。最近的实验证明，肿瘤促进剂烟碱可抑制肿瘤坏死因子。紫外线、化疗药物和 Ca^{2+} 载体介导的程序化细胞死亡，提示程序化细胞死亡可以抑制肿瘤的发生。

某些病毒癌基因介导的细胞转化也涉及程序化细胞死亡，如研究较多的腺病毒基因 E1A 和 E1B。E1A 单独转染原代培养细胞，使细胞增殖率大大增加，如果细胞表达野生型 p53 蛋白，则使程序化细胞死亡大大增加。E1B 可以抑制 p53 介导的程序化细胞死亡，对转化起决定性作用，其分子机制是 E1B 蛋白能与野生型 p53 结合而使其构象向突变型转化，使 p53 的功能丧失，从而在腺病毒转化过程中对转化的启动起负调控作用。

大多数抗癌药物，如拓扑异构酶抑制剂、烷化剂、抗代谢物和激素拮抗剂等，都可在不同类型的肿瘤细胞中诱导程序化细胞死亡。大量的研究表明，抗癌药物与各自特异靶细胞的相互作用，也取决于这些药物诱导程序化细胞死亡的能力。因此，肿瘤细胞对程序化细胞死亡的敏感度是决定化疗效果的关键因素。如何在治疗中提高特异肿瘤细胞死亡增殖比已成为新的化疗目标和筛选抗癌新药的标准。

前列腺癌细胞常常表现雄激素的不依赖性，这是很危险的。但这些细胞仍然可以通过其他信号诱导进入凋亡途径，例如，用一种植物衍生出来的倍半萜 γ 内酯处理这种细胞，可提高并维持细胞内的 Ca^{2+} 浓度，从而诱导细胞凋亡，据此已经研制了一种新药治疗前列腺癌，这种药物只有在前列腺癌组织中有高水平的前列腺特异抗原（PSA）时，其活性才释放出来。

在肝癌诱变剂和促进剂的研究中发现，内菲诺品和苯巴比妥等非诱变致癌物的作用是通过抑制凋亡而促进癌前病灶内细胞存活的，提示人们能够找到保护性的营养物或药物，甚至只要简单地限制摄入的能量总量，就可以促进癌前细胞或癌细胞的程序化死亡。2-氯脱氧腺苷（2-cdA）就是这样一种药物，有效地作用于正在增殖的或不增殖的细胞，引起细胞内 DNA 链断裂和辅酶（NAD）消耗，最终导致程序化细胞死亡。但是在体外用佛波酯处理，可以防止 2-cdA 介导的 DNA 降解，提示佛波酯的处理，激活了蛋白激酶（PKC），而 PKC 干扰了内切核酸酶的激活。

增强抗肿瘤药物特异地诱导程序化细胞死亡或抑制药物对正常细胞的损伤是

改善化疗的有效措施。目前临床上使用的粒细胞集落因子（G-CSF）可诱导细胞分裂，增加细胞存活率，化疗后病人给以 G-CSF，可使成熟及未成熟的粒细胞增加分裂，延长寿命，使化疗对白细胞的损伤很快恢复。现在已知的大多数细胞因子均能抑制抗肿瘤药物诱导的程序化细胞死亡，提示内源性的细胞因子能影响化疗结果。同时，也有文献报道，在某些情况下，提高细胞内 IL-2，IL-3，IL-6 及 G-CSF 的水平，对抑制肿瘤的发生有一定作用。

毫无疑问，随着程序化细胞死亡机制的阐明，人们必将越来越多地把相关的知识应用于各种临床疾病，特别是肿瘤的预防与治疗。虽然目前还不能宣布已经战胜了肿瘤，值得欣慰的是，在攻击肿瘤的武器中又增加了一个极有价值的新武器。

第二节　细胞凋亡的诱导因素、受体、相关基因

一、细胞凋亡的诱导因素

诱导凋亡的因素众多，可分为：①物理因素，如放射线、紫外线、高温；②化学因素，如抗癌药物、氧自由基、氧化氮；③微生物，如 HIV、EB 病毒；④缺乏细胞生长必需因子、如肝细胞生长因子、雌激素（乳腺依赖性）、雄激素（前列腺依赖性）；⑤免疫性，是指抗原和免疫应答过程中所产生的细胞因子和细胞膜分子。另外，慢性缺血、去除血清或生长因子、营养耗尽以及生存信号的缺失也可导致细胞凋亡。本节主要介绍免疫性诱导细胞凋亡的有关因素。

1．细胞因子的作用

不同的细胞因子对于不同的细胞有不同的作用，可诱导细胞凋亡或抑制细胞凋亡，后者常表现为缺乏某种细胞因子而产生“撤退性”细胞凋亡。对同一种细胞最终调控的结果往往表现为多种因子相互拮抗或相互协同的综合作用。

白细胞介素 IL-6 对某些细胞可诱导凋亡，而多数 ILs 起抑制凋亡作用（表 14-1）。

表 14-1　某些白细胞介素在细胞凋亡中的作用

白细胞介素	靶细胞	效应
IL-1	胸腺细胞	抑制凋亡
IL-2	T 细胞	增殖
IL-4	慢性 B 细胞，白血病细胞	抑制凋亡
IL-5	外因血嗜酸性粒细胞	抑制凋亡
IL-6	人早幼粒白血病、细胞系 U937 成熟中性粒细胞	诱导凋亡

2. 造血生长因子

此类细胞因子主要包括 EPO（促红细胞生成素）、G-CSF、GM-CSF、IL-3、SCF（干细胞因子）等。造血生长因子一方面促进造血祖细胞和造血干细胞的存活、增殖、分化和发育；另一方面调控骨髓造血细胞及外周血细胞的寿命，使血细胞的增殖和死亡保持适当的比例，维持血细胞数目在一定范围内保持动态平衡。造血生长因子调控血细胞发生于胚胎期，并贯穿整个造血过程。

3. 肿瘤坏死因子（TNF）

TNF 既可诱导细胞坏死，又可诱导细胞凋亡。细胞的死亡方式与细胞种类有关，如 TNF-α 可抑制体外培养的单核细胞发生凋亡，却可诱导一些肿瘤细胞的凋亡。TNF 作用原理是通过 TNF 受体激活细胞膜的磷脂，产生第二信使神经酰胺。其作用类似 Fas 配体与 Fas 之间的作用。

二、与细胞凋亡相关的受体

1. Fas 和 FasL

Fas 又称 APO-1，现命名为 CD95 分子。Fas 与其配体 FasL（Fas ligand）相互作用是引起细胞凋亡的主要途径之一，因而是当前免疫性凋亡研究的热点。

（1）Fas 属 TNFR/NGFR 家族成员

Fas 基因定位于人染色体 10q24.1，小鼠中是在第 19 号染色体上。其编码产物分子质量为 45kDa I 型跨膜蛋白，由胞外区（157aa）、跨膜区（17aa）及胞浆区（145aa）组成。其胞浆区的 60～70 氨基酸序列与 TNFR 胞浆区有高度同源性，因介导细胞死亡，故称之为死亡域（death domain），因此，Fas 又可称做死亡分子。Fas 主要以膜受体形式存在，还可通过在转录水平的不同拼接产生可溶性 Fas 分子，后者在细胞凋亡中起调节作用。Fas 较广泛地分布于多种类型的细胞，包括外周活化 T、B 淋巴细胞、NK 细胞、单核细胞、成纤维细胞、内皮细胞等。某些组织细胞经活化后也可诱导 Fas 表达。

Fas L 为 TNF 相关的 II 型分子，分子质量 40kDa，其 N 端在胞浆内，跨膜部有一疏水域，其 C 端位于胞膜外，胞外区约 150 个氨基酸与 TNF 家族成员高度同源。因 Fas L 能与 Fas 结合诱发凋亡，故把 Fas L 称死亡因子。鼠 *Fas L* 基因定位于第 1 号染色体。鼠和人 Fas L 蛋白分子同源性高达 76.9%，且无种属特异性，所以人的 *Fas L* 基因转染给小鼠细胞，可诱发凋亡。Fas L 亦可分泌或脱落至细胞外，成为可溶性的功能活性分子。Fas L 主要表达于活化的 T 淋巴细胞。

（2）Fas 和 Fas L 的生物学效应

TNFR/NGFR 超家族中，只有 Fas 和 TNFR1 型可介导细胞凋亡，而且 Fas

介导细胞凋亡必须具备 3 个条件：①细胞表达足够密度的 Fas；②Fas 与其配体或单抗作用形成三聚体；③细胞关闭抗凋亡程序而成为凋亡敏感型细胞。许多组织均可表达 Fas，但仅免疫系统对 Fas 系统介导的细胞凋亡最为敏感。如 Fas 可诱导胸腺细胞阴性选择，诱导外周自身反应性 T 细胞凋亡，诱导被某些抗原活化的 T、B 细胞凋亡以及诱导白血病细胞凋亡。

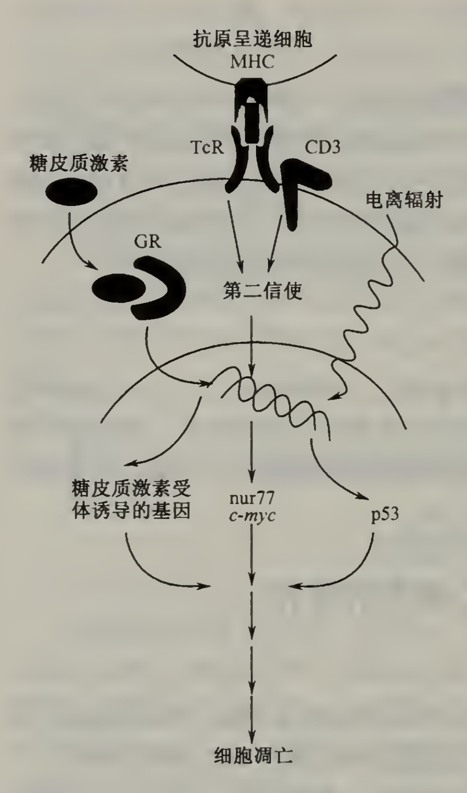


图 14-2　外周 T 细胞凋亡时的信号传递通路

2. T 细胞受体（TCR）

胸腺中未成熟 T 细胞通过 TCR 与自身抗原相结合可发生凋亡（图 14-2，14-3）。CD3 单抗可诱导成熟 T 细胞增殖，对未成熟的 T 细胞则导致凋亡，但凋亡作用可被佛波酯（PMA）所拮抗。在小鼠胸腺未成熟 T 细胞培养结束前 18h 加入 CD3 单抗，观察到细胞数减少了 45%，同时伴有大量 DNA 片段释出。电镜下也只见到典型的凋亡形态特征。用 CD4 单抗预先处理 T 细胞，然后再用 TCR 抗体一起孵育过夜，就可见大量细胞凋亡，而将 CD4 单抗与 TCR 单抗同时加入 T 细胞孵育则不出现凋亡而发生增殖，说明 CD4 分子对细胞在接受刺激时是否发生凋亡也起作用。

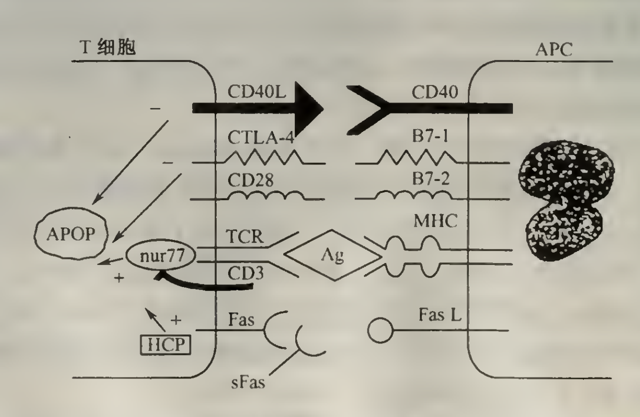


图 14-3　抗原诱导 T 细胞凋亡的机制

三、细胞凋亡相关基因

大多数细胞在凋亡过程中需要新的基因表达，因而人们认为，发生凋亡的细胞是接受外来信号后，发生自杀，而不是他杀。对胸腺细胞和 T 细胞比较研究表明，胸腺细胞经糖皮质激素处理后发生凋亡，而成熟的 T 细胞经糖皮质激素同样处理后不发生凋亡，说明前者表达了自杀基因，而后者没有表达。进入 20 世纪 90 年代，应用分子生物学技术进一步开展了凋亡相关基因和诱导凋亡信号传导途径的研究。对线虫和哺乳动物的研究证明存在凋亡相关基因。一类为促进凋亡基因或称为细胞死亡基因（cell death gene，Ced），线虫为 *Ced-3*，*Ced-4*，哺乳类为白细胞介素-1β 转换酶（IL-1β converting enzyme，ICE）基因，*p53* 基因，*Fas* 基因，*nur77* 基因等。另一类为抑制凋亡基因或称为细胞死亡抑制基因（cell death suppressor gene），线虫为 *Ced-9*，哺乳类为 *bcl-2*（B-cell lymphoma/leukemia-2）和 *Bax* 等。

1. 线虫细胞凋亡的基因调控

人们对细胞凋亡发生机制的认识大多来自对美丽隐杆线虫（*caenorhabditis elegans*，CE）发育中的凋亡基因调控的研究。体积微小的 CE 在胚胎发育期间共产生 1090 个体细胞，而在发育过程中有 131 个细胞死亡，因此，成熟的 CE 有 959 个体细胞。CE 发育中的细胞死亡是一种典型的细胞凋亡。与 CE 死亡有关的基因共有 14 个。各自发挥不同功能，其中直接控制 CE 凋亡的基因主要有 3 个，即 *ced-3*，*ced-4* 和 *ced-9*. 在缺失 *ced-3* 或 *ced-4* 的 CE 突变体中不发生凋亡，而在成长期仍保留多余细胞。Ced-9 对 ced-3 和 *ced-4* 具有调控作用，可防止凋亡过度发生。*Ced-9* 过度表达时，可抑制 *ced-3* 和 *ced-4* 的功能，使凋亡不能发生。而 *ced-9* 功能不良时 CE 往往在胚胎期即因细胞过度凋亡而死亡。因此，*ced-9* 是 CE 发育过程中控制细胞生与死的开关。

2. 哺乳类动物细胞凋亡相关基因

哺乳类动物细胞凋亡的基因调控要比线虫复杂得多。主要有下述 4 类。

（1）*bcl-2* 基因家族

bcl-2 基因家族的基因包括：*bcl-2*，*bcl-x*，*bax*，*mcl-1* 和 *A1* 等。

1）*bcl-2*。*bcl-2* 因首先发现于人类 B 细胞淋巴瘤而得名，随后发现 Bcl-2 可调节细胞凋亡。Bcl-2 蛋白的分子质量为 25kDa～26kDa。*Bcl-2* 转基因小鼠可明显延长小鼠细胞寿命。即使在某些生长因子缺乏时，Bcl-2 也可阻止骨髓细胞和淋巴细胞死亡，并阻止 c-myc 介导的细胞死亡。但 Bcl-2 的表达不能阻止所有类型的生理性细胞死亡，也不能使细胞抵抗 CTL 的杀伤作用。Bcl-2 抑制哺乳类动

物细胞凋亡的能力类似于 Ced-9 抑制 CE 细胞凋亡的能力。人类 *bcl-2* 基因转染 CE 时，也可阻止 CE 中 *Ced-3* 和 *Ced-4* 所介导的细胞凋亡。Bcl-2 的作用广泛，如哺乳类动物细胞对叠氮钠、秋水仙碱、DNA、RNA 和蛋白质合成抑制剂，类固醇，射线作用以及生长因子耗竭所致的死亡的抵抗作用。

2）*bcl-x*。*bcl-x* 编码 241 个氨基酸的蛋白质，与 Bcl-2 有 74%同源性。进一步分析发现，*bcl-x* 分两型，一型为 *bcl-xL*，编码 241 个氨基酸，存在于长寿细胞（如神经元）中，与 bcl-2 有相同作用；另一型为 *bcl-xs*，缺乏 *bcl-2* 编码的 63 个氨基酸的高度保护区，存在于短寿细胞中（如未成熟的 $CD4^+$，$CD8^+$ T 细胞)，对 bcl-2 有抑制作用。

3）*bax*。Bax（bcl-2 associated x protein ）蛋白由 192 个氨基酸组成，分子质量为 21kD。它与 Bcl-2 蛋白有 21%同源性，可与 Bcl-2 形成异二聚体，从而使 Bcl-2 失活，故 Bax 在功能上与 Bcl-2 相反。Bcl-2 和 Bax 的表达程度决定了细胞命运。Bax 占优势时细胞趋向凋亡，而 Bcl-2 占优势时阻止细胞凋亡发生。

4）*mcl-1* 和 *A1* 的生物学功能至今不明。

（2）*ICE* 基因

ICE 即白细胞介素-1β 转换酶，在体内是以非活性形式存在的，在颗粒酶(来自杀伤细胞）的作用下可形成活性 ICE。ICE 是半胱氨酸蛋白酶，能使无活性 IL-1β 前体转成有活性 IL-1β，后者可诱导 *Fas* 基因表达。又由于 ICE 在结构上与线虫 Ced-3 凋亡蛋白有 28%的同源性，故 ICE 本身有类似 Ced-3 直接诱导凋亡的作用。已知 *Bcl*-2 和 *P53* 基因可拮抗 ICE 诱导凋亡作用，故 *Bcl-2* 失活或 *ICE* 基因过度表达时，易诱发细胞凋亡。

（3）*c-myc* 癌基因

目前已知 c-myc 分子有诱导凋亡活性的结构域，但 c-myc 分子正常功能是使 G0/G1 期细胞进入 S 期，促进细胞增殖。进入 S 期的细胞在生长因子作用下，细胞才能进入有效分裂期，如缺乏生长因子时，则 c-myc 激活 *Fas* 基因即诱发细胞凋亡。因此，c-myc 有既可引起细胞增殖，又可引起凋亡的双向作用，主要决定于其他因子对它的影响。

c-myc 基因在许多哺乳类动物肿瘤（如人类乳腺癌、结肠癌、宫颈癌）中呈高表达，且表达程度与瘤细胞体外生长速度相关，表明 *c-myc* 基因与肿瘤增殖有关。但是，在缺乏营养物质或抗体作用下，*c-myc* 高表达的瘤细胞则易发生凋亡。例如高表达 *c-myc* 基因的抗 T 细胞杂交瘤在抗 CD3 单抗作用下易发生凋亡。

（4）*p53* 抑癌基因

抑癌基因 *p53* 在细胞凋亡中起重要作用。射线或化疗药物引起淋巴细胞 DNA 损伤引起的细胞凋亡绝对需要 p53 蛋白的存在，而糖皮质激素，钙离子载体和衰老引起的细胞凋亡却无需 p53 蛋白参与。p53 蛋白功能是多方面的：①与特定序列 DNA 结合，使细胞停滞于 G1 期；②与复制蛋白 A 结合，阻止细胞进

入S期；③是*WAF/cipl*基因的转录因子，诱导产生p21，而后者又与周期素依赖的激酶（cdk）相结合，从而抑制细胞增殖；④如果损伤的DNA在G1期未能修复，则p53直接诱发细胞凋亡，以防止受损的DNA进入S期，产生突变细胞。

临床上许多肿瘤如结肠癌、肺癌常伴有*p53*突变。这样就使肿瘤细胞不但增殖显著，而且凋亡又明显减弱，故肿瘤多呈恶性生长。

第三节　病毒感染与细胞凋亡

一、病毒基因与细胞凋亡

病毒感染细胞早期（吸附和穿透时期）往往能诱导细胞凋亡。其结果是病毒的复制和当代病毒的产生及在宿主细胞中的扩散受到极大限制，由于病毒的复制和繁殖是宿主依赖性的，病毒利用宿主细胞的能量和物质合成病毒基因组和蛋白质，所以，在长期的进化过程中，病毒已形成一种抑制或延迟宿主细胞凋亡的机制，进入宿主后立即表达出的最早期蛋白常常能抑制宿主的凋亡，以完成自身的生命周期并维持繁衍。病毒编码的细胞凋亡抑制蛋白在凋亡抑制中起重要作用。目前已知的病毒基因编码的细胞凋亡抑制蛋白见表14-2。

表14-2　病毒编码的凋亡抑制蛋白

病毒	病毒编码的凋亡抑制蛋白
人腺病毒	E1B-19K，E1B-55K，E3-14.7K
	E3-10.4/14.5K，E3-6.7K
人巨细胞病毒	IE1　IE2　vICA
牛痘病毒	CrmA
痘苗病毒	SPI-2，A53R
黏病毒	MT2　M11L
猴病毒40	Large　T
非洲黄热病病毒	LMW-5-HL（Bcl-2同源物）
	A224L（IAP同源物）
人免疫缺陷性病毒	Tat　Vpr　p17/24
杆状病毒	p35，p49，IAP
人鼻咽癌病毒	BHRF-1，LMP-1，BALF1
人乳头瘤病毒	E6

二、病毒对细胞凋亡的抑制机制及生物学意义

病毒基因编码的抑制细胞凋亡蛋白种类很多，其抑制细胞凋亡的机制不尽相

同，涉及细胞凋亡信号传导途径中的各个环节。①抑制死亡受体在细胞表面的聚集；②抑制死亡受体与死亡配体的结合；③抑制 Bcl-2 家族中促凋亡成员的功能；④抑制 Ca^{2+} 活性。当受到病毒感染时，宿主发生一系列极其复杂的反应。细胞合成一系列蛋白质（包括水解酶和蛋白酶），并启动吞噬细胞、自然杀伤细胞、细胞毒 T 细胞和辅助 T 细胞，以及中性粒细胞和 B 细胞发挥功能，另外，在病毒侵染细胞的部位，细胞还通过干扰素（IFN）通过传导信号的潜能进行反应。可见，宿主对病毒感染有着极其复杂的反应。病毒能够最终通过宿主的内在防护网而完成感染过程。

在宿主对病毒感染所产生的众多反应中，细胞凋亡是宿主抵制病毒感染的有效方式。病毒感染早期的细胞凋亡可抑制病毒 DNA 的复制，病毒基因的表达，因而导致子代病毒量显著减少。

病毒以各种方式抑制这种感染早期的宿主细胞凋亡，才使得病毒能够进行 DNA 复制，蛋白质合成，完成子代病毒装配，完成病毒本身的生命周期。病毒对细胞凋亡的抑制是在进化过程中自然选择的结果。

从病毒和宿主的相互关系中看，很自然地理解有效抑制宿主细胞在病毒感染早期的凋亡对病毒繁衍的重要性，因而也能够理解正是自然选择的压力，使得那些能够有效抑制宿主细胞早期凋亡的病毒得以选择生存。

第四节　细胞凋亡信号传导途径及调控

细胞凋亡的信号途径就是指装配“死亡复合物”的上游信号通路，也就是从细胞接收或感应到死亡刺激信号直至形成“死亡复合物”这之间死亡刺激信号所走过的路程，目前普遍认为在哺乳动物细胞内，存在两条经典的细胞凋亡信号途径。即：①线粒体介导的细胞凋亡信号传导途径，这也是一种内在的信号传导途径，简称线粒体途径；②死亡受体介导的细胞凋亡信号传导途径，简称死亡受体途径；又称为外在的信号传导途径，另外，内质网介导的细胞凋亡信号作为哺乳动物细胞凋亡的一条新的细胞凋亡信号传导途径，正在为人们接受。但每个信号传导途径并不是彼此独立，它们之间存在着交叉。

一、线粒体介导的细胞凋亡信号传导途径

线粒体在细胞凋亡中举足轻重，是通过利用无细胞体系所进行的一系列研究被发现并被证实。在无细胞体系中，受 Bcl-2 抑制的核固缩和 DNA 片段化依赖于线粒体的存在。在 LiuX 等人 1996 年所进行的无细胞体系实验中，加入脱氧腺三磷酸所诱导的 Caspase 激活依赖于细胞抽提过程中细胞色素 c 从线粒体释放这一过程的存在。另外，细胞凋亡的体内和体外实验均证明，在细胞凋亡过程中细

胞色素 c 从线粒体释放，并且这一过程受位于线粒体上的 Bcl-2 抑制。

线粒体途径在细胞受到诱导 DNA 损伤或抑制 DNA 修复的试剂处理时得到启动，此时细胞色素 c 从线粒体释放到细胞质中，细胞色素从线粒体释放到细胞质中后成为凋亡体（apoptosome）的一个基本组成部分，凋亡体由细胞色素 c，凋亡蛋白酶活化因子-1 和 Pro-Caspase-9 组成。凋亡体形成的结果是 Pro-Caspase-9 被激活。被激活的 Caspase-9 再切割和激活下游的 Caspase。Caspase 切割细胞内多种底物，执行细胞凋亡。在线粒体途径中，除了细胞色素 c 外，还有一些其他的凋亡介导物质在凋亡过程中从线粒体释放到细胞质中，如 APF（apoptotic protein activating factor-1，凋亡诱导因子）。有些细胞的线粒体含 Pro-caspase-3，其功能尚在研究中。

由此可见，从细胞接收或感应到引起 DNA 损伤的刺激信号开始，线粒体作为细胞应激反应的感受器感应刺激，释放细胞色素 c 到细胞质，最终诱发细胞凋亡，线粒体途径如图 14-4 所示。

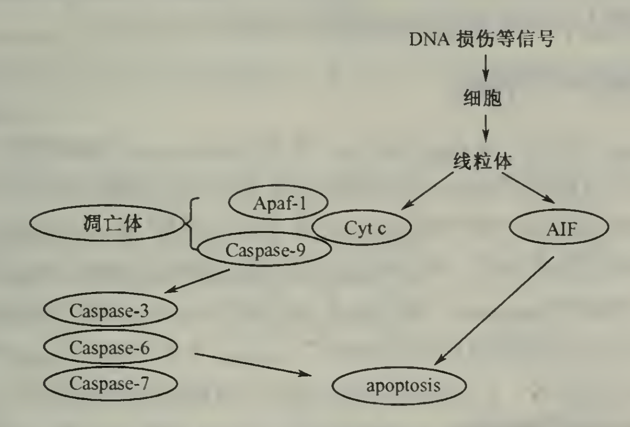


图 14-4　线粒体介导的细胞凋亡信号传导途径

（一）线粒体凋亡信号传导途径的调控

细胞凋亡是一个严密调控的过程。线粒体介导的细胞凋亡传导途径的调控在两个水平上进行。

1. Bcl-2 家族蛋白对线粒体膜完整性的调控

对线粒体膜的完整性的调控主要由 Bcl-2 家族成员完成。Bcl-2 家族成员以不同的方式调控着线粒体膜的完整性，其中抑制凋亡成员能够保持线粒体膜的完整性，抑制凋亡激活因子从线粒体内膜空间的释放。如 Bcl-2 和 Bcl-XL 是 Bcl-2 家族中两个抑制细胞凋亡的成员。二者均可保持线粒体外膜的完整，抑制细胞色素 c 从线粒体释放到细胞质，从而阻断凋亡信号的传导，抑制细胞凋亡。Bcl-2 家族

成员对细胞凋亡的调控作用并不是单向的直线形式，例如 Bcl-2 对细胞凋亡的抑制受 Caspase 的调节，凋亡细胞中活化的 Caspase-3 能够在 ASP34 的位置切割 Bcl-2，Bcl-2 被切割后转移到线粒体中，促使线粒体释放细胞色素 c。从而促进细胞凋亡，可见 Bcl-2 细胞凋亡不仅存在着负调控作用，而且还有正调控作用。

Bcl-2 家族中促进凋亡的成员则破坏线粒体膜的完整性，加速凋亡激活因子从线粒体内膜空间向细胞质释放，促进凋亡发生。例如 BAK 是 Bcl-2 家族成员中的促凋亡蛋白，可通过与线粒体外膜上的电压依赖型离子通道结合，改变线粒体膜透性，使线粒体膜通透性转换孔开关，细胞色素 c 得以释放。Bcl-2 家族中的另一个促凋亡成员同 BID 诱导线粒体释放细胞色素 c 的功能，可被 Caspase-8 的专一性抑制剂 Z-VAD-fmk 及 cr mA 所抑制，进一步的研究发现，BID 只有在被活化的 Caspase 切割以后，才易位到线粒体外膜表面，在线粒体膜上形成孔道，诱导细胞色素 c 的释放。

Bcl-2 家族成员对细胞凋亡的调控作用并不是简单孤立的，它们在对线粒体膜完整性的调控上也同样表现出相互联系和交叉。

2．Caspase-9 激活水平的调控

IAP、Smac/DIALO 和 Htr2A/Omi 均参与 Caspase-9 的激活水平的调控。IAP 家族成员可以抑制 Caspase-9 的激活，而 IAP 的这种抑制作用又可以被另外两个从线粒体释放的蛋白质所抑制。这两个蛋白质即 Smat/DIABLO 和 Htr2A/Omi。最近的研究结果表明：特异性蛋白质-蛋白质相互作用功能区参与 Caspase-9 的激活。人和鼠的 Caspase-9 分别含有 ATPF 和 AVPY 四肽功能区。此功能区与XIAP中的BIR3 功能区相互作用抑制了 Caspase-9 的激活。Smac/DIABLO 结合到 XIAP 的 BTR3 区后，Caspase-9 便从 XIAP 的抑制中释放出来。最近人们发现，Htr2A/Omi 所含的 AVPS 四肽功能区与 XIAP 的相同作用也能够抑制 XIAP 的功能，Htr2A/Omi 与 XIAP、BIR2 和 BIR3 结合也起到促进 Caspase-9 激活从而促进凋亡的功能。

二、死亡受体介导的细胞凋亡信号传导途径

1．死亡受体（death receptor，DR）和死亡配体（death ligand，DL）

死亡受体（DR）是存在于某些细胞表面的一类蛋白质，它们可以通过与死亡配体结合，传导死亡配体携带的凋亡信号迅速激活 Caspase 而诱导细胞凋亡。死亡受体是肿瘤坏死因子（tumor necrosis factor，TNF）受体超家族（superfamily）成员。它们的结构特点是有胞外半胱氨酸富集结构域（cystein-rich domain，CRD）和胞内死亡结构域（death domain，DD）。目前研究最清楚的死亡受体是 CD95（也称 Fas 或 Apol）和 TNFR1（又称 P55 或 CD120a）。已发现的死亡受体

有禽类的 CAR1 和死亡受体 3CDR5（又称 APO_3），死亡受体 4（DR4），死亡受体 5DR5（又称 APO_2，TRAIL-R2，TRICK-2 或 KILLER）。P75 神经生长因子（NGF）受体（又称 NTR）也含有死亡结构域（DD）。P75（NTR）的分子结构和介导细胞凋亡的性能与 CD120a（TNFR1，P55）和 Fas（CD95）有明显的相似性，也属于 TNF/NGF 受体成员。目前已知的死亡配体有 CD95 配体，又称 CD95L（CD95，Fas ligand）其可结合 CD95（即 Fas）；TNF 和淋巴毒素 α 可以结合 THFR1，APO_3 配体（即 APO_3L 或 TWEAK）可以结合死亡受体 3（DR3）；APO_2（即 DR5，TRAI-R2）配体（又称 APO_2L 或 TRAIL）可以结合死亡受体 4（DR4）和死亡受体 5（DR5）。CAR1 的配体尚未发现。

死亡受体途径的主要环节如下：死亡配体携带着凋亡信号，通过与死亡受体结合将凋亡信号传到死亡受体。死亡受体在与死亡配体的结合中寡聚化并被激活。激活的寡聚化死亡受体通过死亡结构域（DD）与接头分子结合，将凋亡信号传到接头分子并使接头分子寡聚化而被激活，接头分子再通过死亡效应结构域（DED）或 Caspase 已接收到凋亡信号，细胞凋亡得到启动。Caspase 在凋亡体中被激活，激活后的 Caspase 或继续激活下游 Caspase，或直接分解细胞蛋白质，最终完成细胞凋亡。

由于死亡受体和接头分子种类较多，其相应激活的 Caspase 又有一定的特异性，所以上述的死亡受体途径中的主要环节，对不同的死亡受体和接头分子而言，每一环节都各具特点。现以 Fas 和 FADD 介导的凋亡信号传导途径为例，说明死亡受体介导的细胞凋亡信号传导途径的分子结构基础。

Fas 的死亡结构域 DD 含有 6 个由短的 loop 环相连的 α 螺旋，其中 α_2，α_2-$\alpha_3$loop 环和 α_3 参与 Fas 和 FADD 死亡结构域 DD 的结合。FADD 的 DD 含有类似的 6 个 α 螺旋。FADD 和 Fas 的 DD 中的 α_2 和 α_3 均含有许多带电荷氨基酸残基，故推测二者的相互作用涉及静电相互作用。

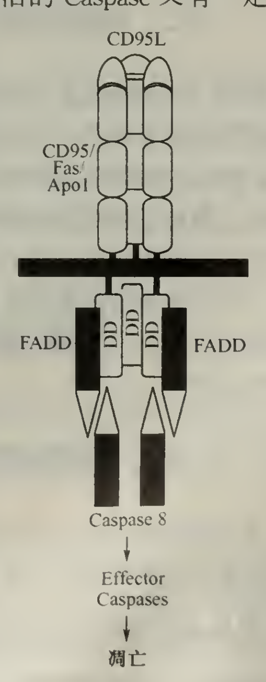


图 14-5　Fas 和 FADD 介导的细胞凋亡信号传导途径

FADD 的死亡效应结构域 DED 具有与其死亡结构域 DD 非常相似的三维结构，而且也含 6 个 α 螺旋。但是螺旋的长度、方向以及连结各螺旋的 loop 环的长度与 DD 的不同。另外，DED 有 2 个疏水区是 DD 所没有的。已有实验结果表明 DD 和 DED 用不同的氨基酸残基进行蛋白质相互作用。DED 与 Caspase 的相互作用并不涉及静电作用，而是通过 α_2 和 α_5 中所含的疏水氨基酸的疏

水相互作用来完成。

由 Fas 和 FADD 介导的细胞凋亡信号传导途径见图 14-5。

2. 死亡受体途径的调控。

死亡受体介导的细胞凋亡信号传导途径可在多个水平上受到调控。

（1）基因表达水平的调控

死亡配体（如 FasL 和 TRAIL）的表达是被调控的。如 Fas 配体在 T 细胞中的表达受 T 细胞激活的上调。死亡受体的表达也是被调控的，如 Fas 的表达随着 B 细胞或 T 细胞激活而增加。

（2）胞外调节分子对死亡受体的调控

FLIP（Flice inhibitory protein）是调控死亡受体介导的细胞凋亡途径的主要调节分子。FLIP 有两个嵌合的 DED，但没有功能性蛋白酶结构域。它通过与 FADD 结合抑制 Caspase 的招募和激活。

（3）黏附分子对死亡受体与死亡配体相互作用的调控

位于 T 细胞上的 FasL 和位于 B 细胞上的 Fas 可以相互作用，从而诱导 B 细胞 Fas 依赖型细胞凋亡。这种特异的 Fasl 和 Fas 的相互作用需有黏附分子的存在。

三、内质网介导的细胞凋亡信号传导途径

内质网（endoplasmic reticulum，ER）是蛋白质翻译、加工的场所，同时又是细胞内钙的储存地之一。当内质网蛋白质过度积累，或内质网中钙离子平衡被打破时，导致内质网应激和细胞凋亡。缺失 Caspase-12 的小鼠对内质网引起的凋亡不敏感，而对其他途径的细胞凋亡依然敏感，故此人们推断，Caspase-12 参与内质网介导的细胞凋亡。同时，由于 Caspase-12 缺失的小鼠发育正常，无任何明显缺陷，所以推测由内质网介导的细胞凋亡途径在某些病理条件下发挥作用。

内质网介导的细胞凋亡是一条新发现的凋亡途径，其具体信号传导途径和机制尚待深入研究。

四、细胞凋亡信号传导途径的相互交叉

上述三条细胞凋亡途径并不是绝对孤立的，它们之间在一定条件下互相交叉，共同调节细胞凋亡这一重要生命过程，使之利于生物个体的生存。

1. 死亡受体途径与线粒体途径间的相互交叉

Caspase-8 在死亡受体途径中被激活。激活的 Caspase-8 可切割 Bcl-2 家族成员 BID 形成截短的小 BID。截短的小 BID 转位到线粒体上，改变线粒体膜的通透性，促进细胞色素 c 从线粒体向细胞质释放，从而激活线粒体介导的细胞凋亡

途径。由此，死亡受体途径通过 BID 激活了线粒体途径。这两条凋亡途径的相互交叉见图 14-6 所示。

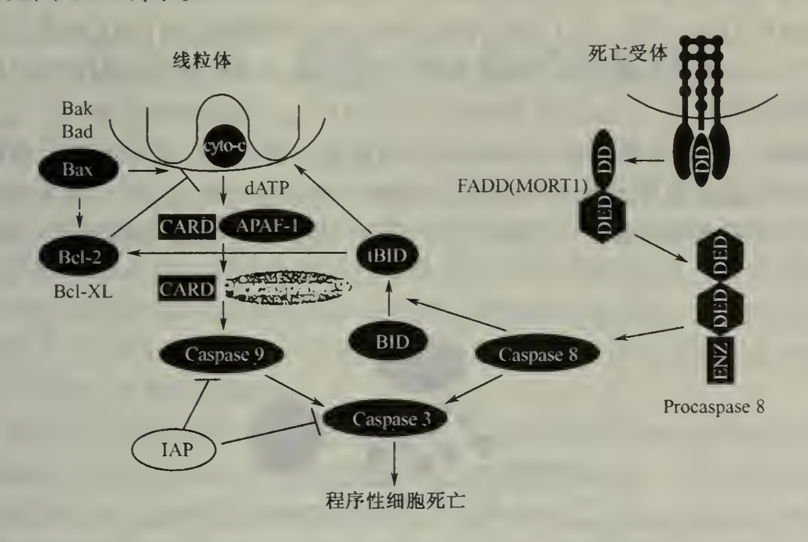


图 14-6　BID 是线粒体凋亡途径和死亡受体凋亡途径交叉的媒介

2. 线粒体途径和内质网途径的交叉

已有证据显示，线粒体途径和内质网途径间存在凋亡信号的相互对话(cross-talk)，当内质网新的 Ca^{2+} 释放时，与内质网邻近的线粒体感受到的钙离子浓度比胞浆中平均钙离子浓度高得多，这使线粒体摄入钙离子能力增加，导致钙离子超载，线粒体外膜破裂，线粒体内的凋亡相关因子细胞色素 c、A1F1 释放到细胞质中，启动线粒体介导的凋亡途径。

五、CTL 诱导靶细胞的凋亡

近几年来，对 CTL 杀死靶细胞机制的研究引起了免疫学家的极大兴趣。CTL 强大而特异的溶细胞活性至少可通过两条途径来实现。即穿孔素（perforin）介导和 Fas 介导。

1. 穿孔素介导的靶细胞死亡

当 CTL 的 TCR 与靶细胞上 MHC 呈递的抗原结合时，CTL 排出胞浆颗粒，后者含有穿孔素、颗粒酶等成分。颗粒酶包括颗粒酶 A（granzyme A，GzmA）和颗粒酶 B（granzyme B)，颗粒酶可通过穿孔素在靶细胞膜局部形成的孔道或通过靶细胞内吞作用进入靶细胞。GzmA 切割 IGAAD（inhibited granzyme A activated dnase 即 set)，使后者不再结合并抑制 GAAD（NM23-1)。未被抑制的

GAAD 催化 DNA 单链切口，GZMB 直接或间接通过 Caspase 损伤 CAD 的抑制剂 ICAD。被切割的 ICAD 可与 CAD 结合，转移到核内切割双链 DNA 产生大小相差 180bp 的 DNA 片段。由 GZMA 催化形成 DNA 单链切口的 DNA 链大片段，可以加快 GzmB 催化形成寡核小体的速度。所以 GzmA 和 GzmB 在促进染色体分裂上是呈现协同作用。GzmB 诱导的凋亡途径中有 cytc 的释放和依赖 Pro Caspase9/3 的激活等，故对于有高表达 Bcl-2 的肿瘤细胞，对 GzmB 诱导的凋亡表现抵抗性，但对 GzmA 诱导的凋亡作用呈现敏感性。在 GzmA 和 GzmB 二者的诱导凋亡途径中，已知尚有许多正负性调节因子参与，关于它们的调节机制正在研究中。如图 14-7。

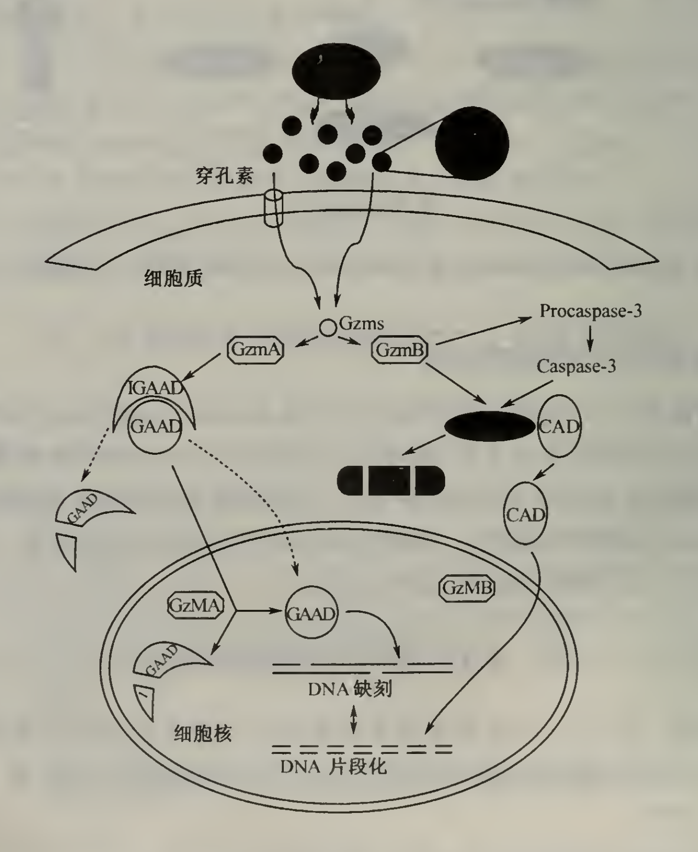


图 14-7　CTL 介导的凋亡 DNA 剪切的模型图

2. Fas 介导的靶细胞凋亡

聚合穿孔素的形成，有利于颗粒酶等成分进入靶细胞，进而使 ICE 活化，通过 ICE 途径诱导细胞凋亡。由于缺失穿孔素的 CTL 仍能杀伤 Fas^+ 的靶细胞，但都不杀伤 Fas^- 的靶细胞，表明 FasL 与 Fas 结合是 CTL 杀死靶细胞的另一重要机制。

第十五章　重组 DNA 技术及其应用

重组 DNA 技术（recombinant DNA technique）又称基因工程（gene engineering）或基因操作（gene manipulation），是指将来源于不同生物或者人工合成的 DNA，有目的的在体外进行改造和重新组合，再转入微生物或真核细胞内进行无性繁殖（cloning），使重组基因在细胞或生物体内得到表达，从而改变生物的遗传特性或者创造新类型的生物。与经典的遗传学手段相比，重组 DNA 技术是在基因水平上进行的操作，目的更明确、方法更简捷，因而自发展之初它就成为遗传工程中一个重要的领域。

重组 DNA 技术兴起于 20 世纪 70 年代，它的发现有深刻的理论依据和坚实的实验基础。20 世纪 40 年代 Avery 等证实 DNA 是遗传物质、50 年代 Watson 和 Crick 提出的 DNA 结构的双螺旋模型以及 60 年代以 Nirenberg 为首的科学家发现的 64 个遗传密码和遗传中心法则，为重组 DNA 技术提供了理论根据；20 世纪 60 年代末，限制性内切核酸酶、DNA 连接酶、载体的发现则为重组 DNA 技术奠定了实验基础：限制性内切核酸酶使分离单个基因成为可能，DNA 连接酶使体外基因重组得以实现，而载体则提供了单个基因富集以及基因表达的途径。

随着分子生物学的发展，重组 DNA 技术现已成为一门内容广泛的综合性生物技术学科，为基础理论研究和应用科学带来了革命性的进步。在理论研究方面，重组 DNA 技术可以分离和扩增到某一具体基因，从而可以研究基因的精细结构和功能。人们借助重组 DNA 技术，不仅明确了多种基因的结构及功能，还发现了断裂基因、重叠基因、假基因等多种基因类型，同时在基因表达调控方面也取得了很多本质上的认识，所有这些，使人类对生命现象的认识在分子水平上产生了质的飞跃。在生产应用方面，人们利用重组 DNA 技术，使基因在微生物、真核细胞乃至动植物中表达，获得了许多重要的产品，尤其是医药领域，如干扰素、激素、免疫调节剂和蛋白质疫苗等；利用重组 DNA 技术，人们还创造出新的抗病、抗虫害、优质高产的动植物品种或生物反应器，到目前重组 DNA 技术已经为人类带来了巨大的经济效益和社会效益。

重组 DNA 技术包含三个要素：载体、工具酶和表达系统（原核或真核宿主细胞）。其技术路线大致包括几个程序：①目的 DNA 片段的分离制备；②目的片段与载体的体外重组；③将重组的 DNA 分子转化到表达系统中（原核或真核宿主细胞）；④筛选、鉴定重组 DNA 分子。下面我们围绕这些要素和程序分别进行叙述。

第一节　载体和工具酶

一、载　　体

载体（vector）是指能将目的DNA片段带入宿主细胞、并能进行扩增的一类DNA分子。载体一般都能在细胞内建立稳定的遗传状态，可在细胞内独立繁殖、传代，或是整合到宿主细胞染色体基因组中随基因组的复制而复制。表达型载体还可表达携带的目的基因。载体应具备以下条件。①自主复制：载体必须含有复制起点，可在适当的宿主细胞中自主复制。②包含多克隆位点：载体上应具有多种限制性内切核酸酶的位点，便于外源目的基因的插入和连接。③具有可供选择的遗传学标记：借助这些标记可以区别转化和未转化的宿主细胞、阳性重组分子和未重组分子。④其特征是被充分掌握的：应知道载体的全核苷酸序列；对于载体基因的作用、载体基因之间的相互关系以及载体基因与宿主细胞的基因或蛋白之间的相互关系，应该有深入的了解。需特别注意的是，载体必须是安全的，在任何对于人类和其他物种的危害程度未明了之前，载体的使用应置于严格控制之下。

一般而言，载体是通过改造天然质粒、噬菌体和病毒而成。目前，全世界构建和应用的载体有数千种之多。根据载体用途、应用对象不同，可以对其分类。从用途可以将载体分为克隆载体、表达载体、转移载体、探针载体等；从应用对象可以分为原核克隆载体、植物克隆载体、动物克隆载体。

（一）质粒载体

质粒（plasmid）是存在于细菌染色体以外的独立的复制子。它们是一类闭合环状的双链DNA分子，不同质粒分子大小差别很大，从1.5kb至100kb以上都有，多数质粒大小在3～10kb之间。在革兰氏阴性细菌中，质粒的复制有两种模式：严紧型和松弛型，严紧型质粒的复制除了依赖于宿主的DNA聚合酶III外，还依赖于宿主其他蛋白质，通常在一个细胞中只有1～5个拷贝，这对于基因的扩增和表达是不利的；松弛型质粒的复制则不同，除了DNA聚合酶III，基本上不依赖宿主其他蛋白质。因此，即使在宿主染色体复制和蛋白质合成停止时，仍能大量扩增。通常一个细胞中可含30～50个松弛型质粒，大量扩增时甚至可达上千个。松弛型质粒在重组DNA技术中得到广泛应用。

用于重组DNA技术的质粒载体都是人工重建的，本身的DNA大小多在3kb～5kb间，含一个复制原点，一个多克隆位点区，一个或两个抗药基因以及用作遗传标记的报告基因。抗药基因和报告基因用于下一步的筛选。质粒进入细

胞是通过转化技术实现的，但重组质粒大于 15kb 时，用普通的冷 $CaCl_2$ 转化技术转化时效率明显降低，因此质粒一般用于 10kb 以下的 DNA 片段的克隆。下面是几种代表性的质粒载体。

1.pBR322 质粒载体和 pUC 系列质粒载体

重组 DNA 技术早期发展中，pBR322 质粒是应用得最为广泛的质粒克隆载体之一。pBR322 质粒分子大小 4.3kb，内含松弛型质粒 C_olE1 的复制原点、选择性标记基因 Ap^r（抗氨苄青霉素基因）和 Tc^r（抗四环素标记基因）。在 Ap^r 基因和 Tc^r 基因内各有一些限制性内切核酸酶的位点（图 15-1），一旦被插入外源片段，这个基因即被灭活，产生的重组质粒所转化的细菌不再具备对该抗生素的抗性，利用这种性质就很容易筛选所需的重组质粒。pBR322 还常常用作构建新载体的基础，如 pBR324～329 等一系列的克隆载体都是从 pBR322 出发构建的。

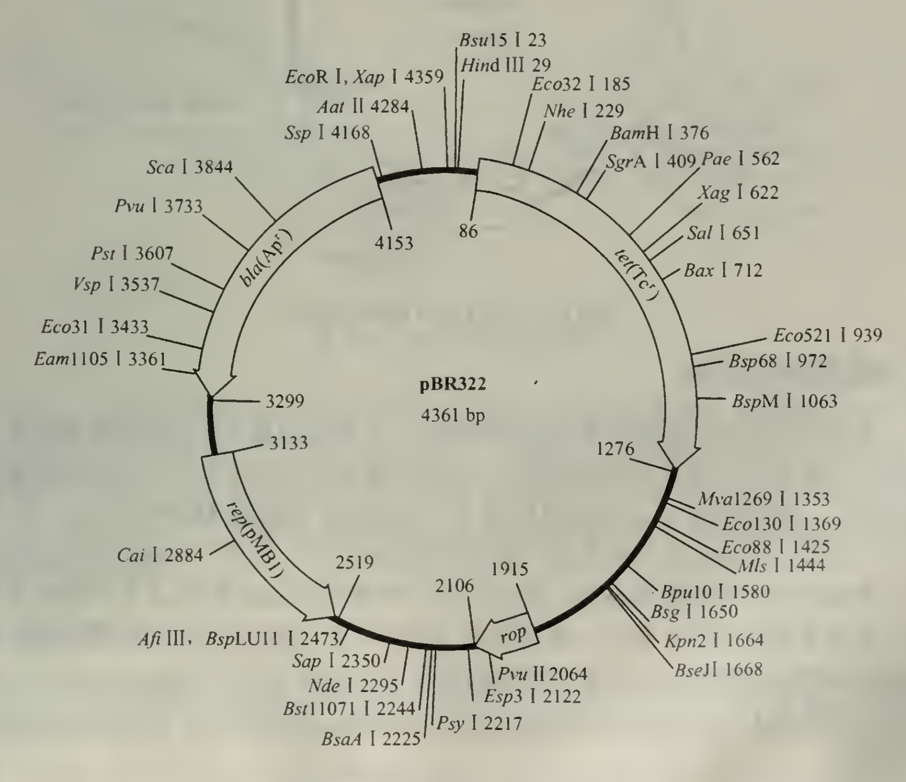


图 15-1　pBR322 质粒示意图

pUC 系列质粒载体也是常用的质粒克隆载体，由 pBR322 载体衍生构建而成。它是用 *E.coli* 的乳糖操纵子基因 N 端部分（*lacZ′*）代替了 pBR322 的选择

标记 Tc^r 基因（图 15-2），并在 *lacZ*′中组装了多克隆位点区。当这种质粒被转化到可以与 *lacZ*′基因互补的大肠杆菌中，在含 IPTG（异丙醇-β-D 硫代半乳糖苷）和 X-gal（5-溴-4-氯-3-吲哚-β-D-半乳糖）的诱导培养基上培养时，可形成蓝色菌落；如果在 *lacZ*′中的多克隆位点区插入外源片段，可造成 *lacZ*′失活，形成白色菌落。这样可以直接筛选重组 DNA，而不必采用类似于 pBR322 质粒相对麻烦的负筛选法进行筛选。

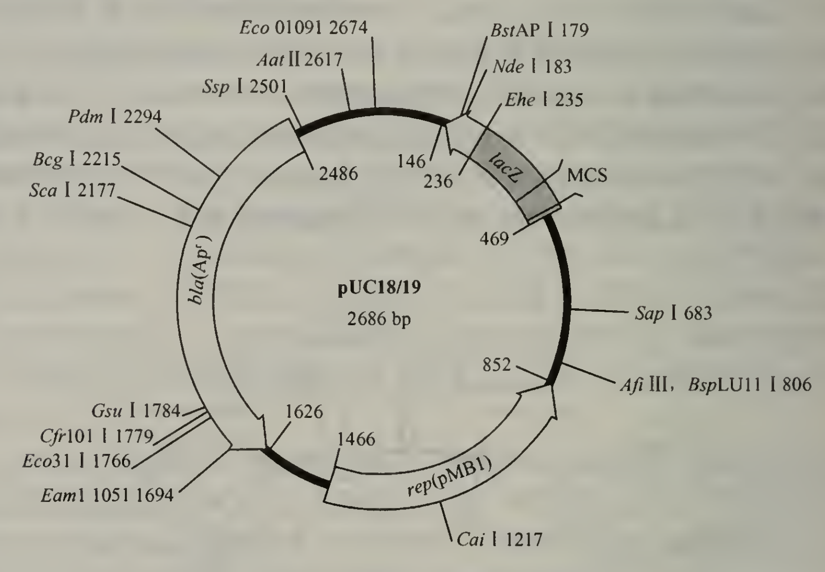


图 15-2　pUC18/19 质粒示意图

2. 细菌质粒表达载体

通常用于表达外源目的基因的质粒载体，主要目的是为了在细菌中表达真核基因，因此这类载体有个共同的特点：在多克隆位点区的上游有强大的原核基因启动子（一般来源于 *lac*、*trp* 或使用 T7 promoter）和细菌核糖体结合的 SD 序列，在克隆位点的下游常常还有一段转录终止的序列。这样可以在转录和翻译水平上保证外源片段的高效表达。图 15-3 是一种强大的表达质粒 pET 质粒的示意图，在多克隆位点的上游有一段天然的 T7RNA 聚合酶启动子。用 pET 质粒在大肠杆菌中表达外源基因，表达的蛋白量常可达细菌总蛋白量的 30% 以上（随基因不同而异），并且可以利用融合的 6 个组氨酸，很方便地用 Ni 柱纯化所表达的蛋白。

3. 体外转录质粒载体

由 pUC 质粒构建而来，主要有 pGEM 质粒系列和 pSP 质粒系列。pGEM 质粒含有一个 Ap^r 基因和一个 *lacZ*′基因，多克隆位点区位于 *lacZ*′基因内（图

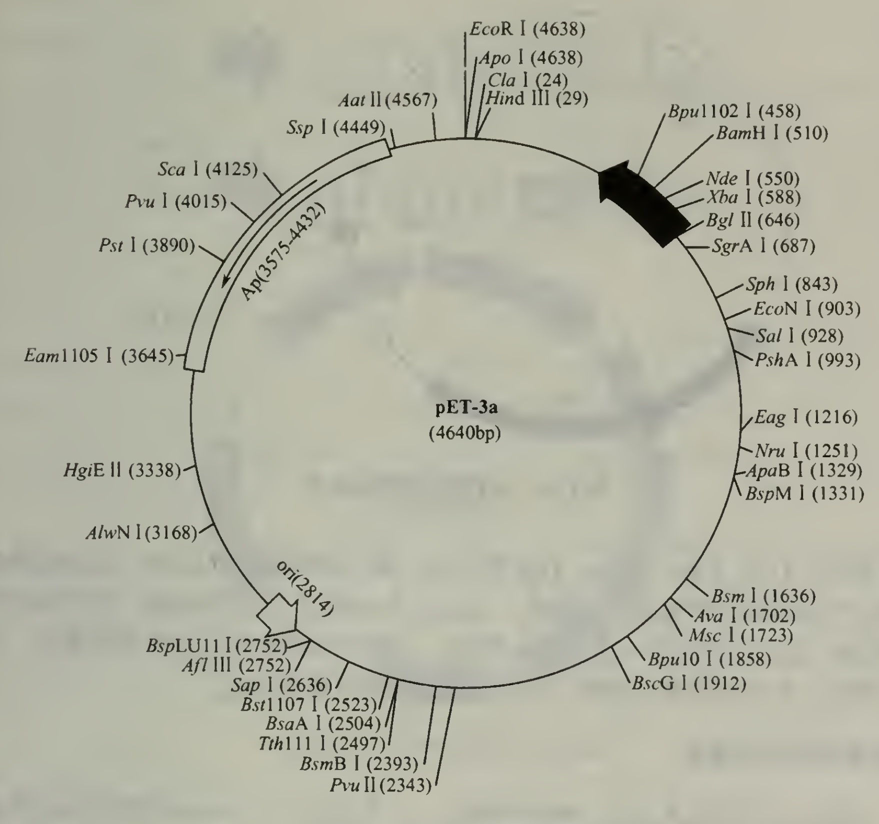


图 15-3　pET-3a 质粒示意图

15-4)。pGEM 质粒与 pUC 质粒主要的不同在于 pGEM 质粒的多克隆位点的上、下游分别含有噬菌体 T7 和 SP6 的启动子，在体外可以利用 T7 和 SP6 的 RNA 聚合酶进行转录。pSP 质粒利用的是噬菌体的 SP6 启动子。体外转录质粒提供的 mRNA 转录本可以用作制备 RNA 探针，也可用于兔网织红细胞裂解物翻译系统中进行体外翻译。

4. 穿梭质粒载体

穿梭质粒载体（shuttle plasmid vector）是指含有两种不同的复制起点和选择标记的重组质粒，它可以在两种不同的宿主细胞中存在和复制。这种特性可以使质粒穿梭存在于真核和原核两种细胞之间，对于重组 DNA 技术的研究非常有用。

pKSV-10 是一种细菌-哺乳动物细胞穿梭质粒载体，由 pBR322 的一段 DNA 和病毒 SV-40 的一段 DNA 组建而成。它具有 pBR322 和 SV-40 的两种复制原点，既能在细菌宿主中复制和扩增，也可以在真核宿主中复制和表达。

YEp24 是一种细菌-酵母穿梭质粒载体。根据复制模式不同可以将酵母质粒

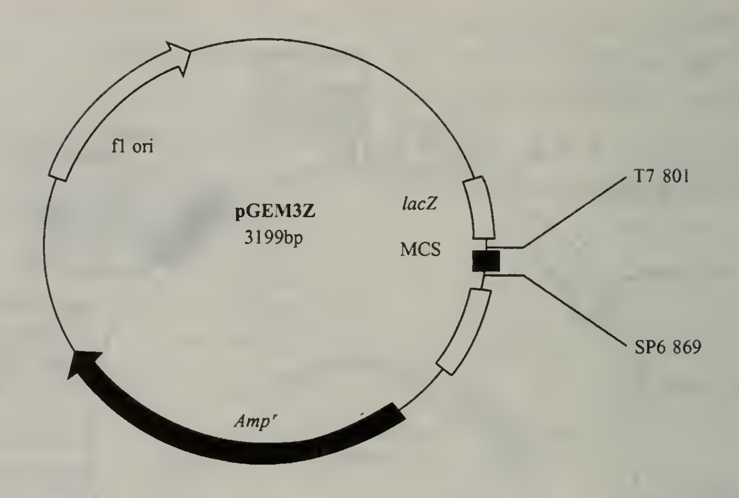


图 15-4 pGEM3Z 质粒示意图

分为五类，YIp、YRp、YEp、YCp 和 YLp，除了线性质粒 YLp 外，其余的质粒均能与大肠杆菌质粒构成穿梭载体。YEp24 是由酵母 2μm 质粒的一部分加上酵母染色体 *URA3* 基因的一部分，分别插入 pBR322 的两个不同位点而成的。因此，既可以在大肠杆菌中复制，也可以在酵母中复制和表达。

5. 真核细胞表达质粒

真核表达质粒实质上是一种穿梭质粒，广泛应用于人和动物细胞中表达外源基因，也可以用于 DNA 疫苗的研究。如 pcDNA 质粒系列（图 15-5），除了含有必需的原核系统元件外，它还含有一个极强的人巨细胞病毒（CMV）立即早期启动子，用于高水平的外源基因表达，并且 pcDNA 质粒可以转染到绝大多数的哺乳动物培养细胞，应用极为广泛。

6. 启动子探针载体

启动子探针载体可用于研究调控的 DNA 序列的特征。它们含有一个产量容易定量检测的报告基因，此基因上游因缺乏启动子而处于非转录状态。常用的报告基因有 β 半乳糖苷酶基因（*Z* 基因)、氯霉素乙酰基转移酶基因（*CAT* 基因)、荧光素酶或者绿色荧光蛋白（*GFP*）基因。一旦在非转录的报告基因上游插入具有启动子活性的 DNA 序列，就可立即激活报告基因进行转录和表达，启动子的活性和强度可以通过定量检测基因产物来确定，除此以外，启动子探针载体还常常用于研究某种蛋白对某一基因表达的影响。

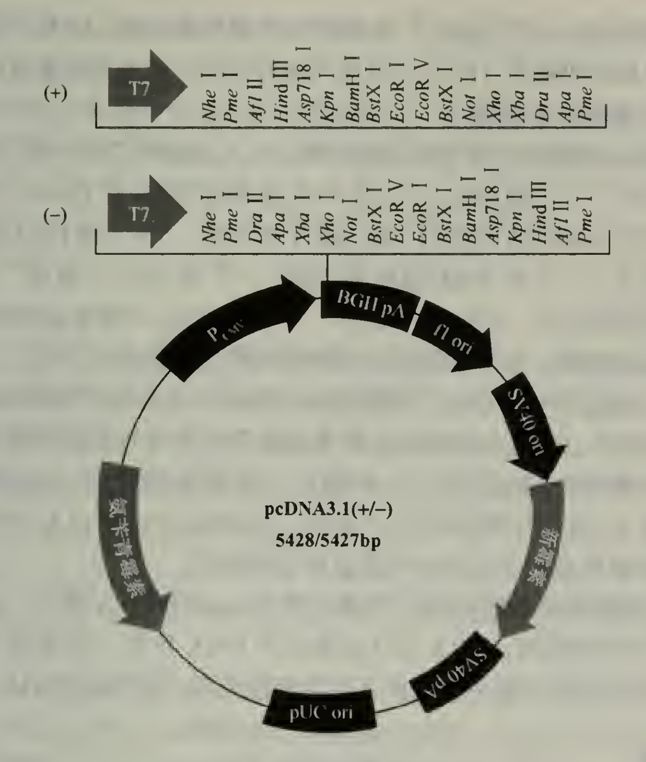


图 15-5 pcDNA 质粒系列示意图

（二）噬菌体载体

噬菌体指专门感染细菌的病毒，由它构建的载体称为噬菌体载体。噬菌体的结构比质粒要复杂得多，就两者的 DNA 本质比较而言，质粒仅仅是一种含有复制原点的裸露的 DNA 分子，而噬菌体 DNA 分子除了含有复制原点外，还含有编码外壳蛋白质、与 DNA 整合、重组等有关的基因以及一些调节基因。从噬菌体（λ 噬菌体）出发构建的载体可以克隆的 DNA 片段比质粒载体可克隆的 DNA 片段大得多，并且重组 DNA 被包装成噬菌体后可通过感染而导入细胞，其效率比质粒 DNA 转化效率要高得多。因此噬菌体载体特别适用于真核基因的研究 (真核基因一般都比较大)。迄今为止，运用噬菌体载体实现了大量成功的基因克隆和表达，表明它们是一种良好的基因克隆载体。目前已有大量的噬菌体突变体供研究者使用，主要用于 cDNA 文库以及表达肽库（噬菌体表面展示）的构建。

1. 单链丝状噬菌体载体（M13 载体）

单链丝状噬菌体载体一般由 M13 单链噬菌体衍生而来。M13 噬菌体是一种

丝状噬菌体，基因组为一有 6407 个核苷酸的单链环状 DNA 分子。基因组上有一段长 507nt 的基因间隔区（IS 区）可插入选择性标记和多克隆位点，其余部分都是噬菌体生长繁殖所必需的。

M13 噬菌体仅感染含 F 因子的大肠杆菌。进入细胞的“+”链 DNA 作为模板复制出互补的“-”链 DNA，由此产生复制型的双链 DNA（RF-DNA）。RFDNA可按环状 DNA 的方式复制，到 200 个拷贝时，“+”链 DNA 可被单链结合蛋白结合，终止“-”链 DNA 的复制，而“-”链 DNA“复制”产生“+”链 DNA 的过程持续进行。这样新合成的“+”链 DNA 不断地被噬菌体衣壳蛋白包装形成新的噬菌体，以非溶菌的方式释放到宿主细胞外。

M13DNA 虽然是单链 DNA，但在细胞内仍可形成环状双链 DNA（复制型 DNA），并且无论单、双链 M13DNA，不需要体外包装均可直接转染大肠杆菌受体细胞，这一点使得 M13 载体成为一种操作十分简便的载体。M13 载体的多克隆位点下游含有 β 半乳糖苷酶基因（*Z* 基因），其筛选方法与 pUC 质粒系列是一样的，通过噬菌斑是蓝色还是白色来鉴定重组噬菌体。

重组 M13 噬菌体的单链 DNA，主要用于 Sanger 法 DNA 测序，也经常用作 oligo 诱导定点突变的模板。M13 可以克隆大的 DNA 片段，但克隆了大片段的 M13 载体在传代过程中不稳定，因此其克隆片段长度一般不超过 1kb。

2. λ 噬菌体载体

λ 噬菌体颗粒中的 DNA 是一线性双链 DNA 分子，长 48 052bp，两端各有 12 个碱基的 5′凸出黏性末端，两个黏性末端是互补的。λDNA 上共有 50 多个基因，几乎所有基因的功能及其表达调控已经研究得非常清楚。λDNA 上的有些区域对于 λ 噬菌体的生长繁殖不是绝对需要的，可以被外源 DNA 片段所取代，这一点是 λ 噬菌体成为克隆载体的基础。

λ 噬菌体感染大肠杆菌后，进入细胞的 λDNA 通过两端的黏性末端环化，在感染早期以环状 DNA 进行转录，随后进入两条不同的生长途径：其一为溶菌循环，环状 DNA 经复制、切割产生子代线性 DNA 分子，与大量合成的噬菌体结构蛋白一起装配成子代噬菌体颗粒，通过宿主细胞裂解释放出子代噬菌体颗粒，感染新的细胞；其二为溶源生长，λDNA 可以整合到宿主细胞基因组中，以无活性的原噬菌体形式随宿主基因组的复制而复制。当宿主细胞处于一定的胁迫条件下时，原噬菌体可以脱离宿主细胞基因组，重新进入循环。

λ 噬菌体载体主要有如下特点：①筛选简便，如果没有外源 DNA 片段的插入，载体本身的 DNA 太小难以被包装，转化后产生的空斑一般都是重组体。此外，还可以利用特殊的 DNA 片段取代 λDNA 的非必要区段，通过改变 λ 噬菌体的表型（如 Spi^- 筛选法）或引入报告基因（如 *LacZ* 基因筛选法）来筛选重组 λ 噬菌体载体。②可克隆的片段大，最大可达 22kb。③转化效率高，重组的 λDNA

都要经过体外包装，成为有感染力的噬菌体，与重组质粒 DNA 转化相比，具有高得多的转化效率。

λ 噬菌体载体是最先使用的分子克隆载体，主要用于 cDNA 文库构建，也经常用于外源目的基因的克隆。λ 噬菌体载体的种类包括 Charon 系列、λgt 系列等。

（三）cosmid 克隆载体

cosmid 克隆载体由质粒载体和 λ 噬菌体的两个黏性末端（cos 位点）组装而成。cosmid 是英文 cos site-carrying plasmid 缩写而成，因此 cosmid 克隆载体又称黏性末端质粒，它集中了质粒载体和 λ 噬菌体载体的优点。载体本身大小 4～6kb，可克隆的 DNA 片段大小可达 30～45kb，广泛应用于基因组 DNA 文库的构建。

（四）植物、动物病毒载体

随着人们对植物病毒和动物病毒的分子生物学机理的深入了解，一些遗传背景研究较为清楚的植物、动物病毒已被构建成载体，用于外源基因（主要是真核基因）的克隆和表达。这些非噬菌体病毒载体的优点显而易见：在真核生物宿主或宿主细胞中表达的真核生物蛋白质，比原核系统中表达的蛋白质在修饰、构象等方面更趋近天然蛋白，因而在活性上更趋同于天然蛋白。

植物、动物病毒载体构建的基本策略大致有几种：①构建转移载体或置换型克隆载体。选择病毒基因组中对其自身生长繁殖非必需的区域，用外源基因进行置换，转染宿主细胞或动植物体进行表达。通常外源基因上游需要设计一个适应于宿主的强启动子以利于基因的高效转录和表达。②构建互补克隆载体。利用两种不同缺陷型病毒之间的互补，表达克隆于其中一种缺陷病毒上的外源基因。由于这一策略可能会重组产生野生型病毒，很少被采用。③构建混合克隆载体。通常这种克隆载体通过质粒与病毒之间的重组来构建，前述的穿梭质粒载体即属此类。

植物病毒中的花椰菜花叶病毒（CaMV）、烟草花叶病毒（TMV）以及动物病毒中的猴空泡病毒 40（SV40）、腺病毒（Ad）、痘病毒等已被广泛地构建成不同类型的载体。值得一提的是，来源于两种杆状病毒：苜蓿银纹夜蛾核型多角体病毒（AcNPV）和家蚕核型多角体病毒（BmNPV）的 AcNPV 表达载体和 BmNPV表达载体，在昆虫细胞中已经成功表达了多种真核蛋白分子，因其操作简便、表达量高、所表达蛋白质活性好而成为目前最好的表达系统之一。此外，目前发展较快的还有反转录病毒载体，其特点是接近 100％的基因转移效率以及具有整合到宿主基因组的能力。

二、工 具 酶

工具酶指在重组 DNA 技术中用于切割、连接、修饰 DNA 或 RNA 的一系列酶。它们是重组 DNA 技术中的基本工具，其中最重要的是限制性内切核酸酶和 DNA 连接酶，是 DNA 重组技术出现的酶学基础。其他常用的工具酶包括 DNA 多聚酶、反转录酶、核酸酶 H、碱性磷酸酶等。本节扼要介绍主要工具酶的性质、用途和使用特点。

（一）限制性内切核酸酶（restriction endonuclease）

1. 种类及命名

限制性内切核酸酶是一类能特异地识别双链 DNA 序列并进行切割的酶，根据酶性质的不同可以分为三类：I 型酶、II 型酶和 III 型酶（表 15-1），由表中可见在重组 DNA 技术中真正有用的是 II 型酶，它既能识别特异性的序列，又能在识别序列中或近旁切割。所以，如不特别说明，重组 DNA 技术中提到的限制性内切核酸酶是指 II 型酶。

表 15-1 不同类型限制性内切核酸酶的主要性质

类型	I 型	II 型	III 型
识别序列	特异，但无规律	特异，旋转对称	特异，但无规律
切割位点	在识别序列前后 100～1000bp 范围内	在识别序列之中或在附近	在识别序列之后 25～27bp 处
切割位点特异性	非特异	特异	特异
所需辅因子	Mg^{2+}、ATP、S-腺苷甲硫氨酸	Mg^{2+}	Mg^{2+}、ATP、(S-腺苷甲硫氨酸)

限制性内切核酸酶主要来源于细菌，在少数霉菌和蓝藻中也有发现。它的命名一般是取来源生物属名的第一个字母（大写）、种名的前两个字母（小写），再加上菌株代号的第一个字母组合而成。如果同一种生物中具有不同特异性的限制性内切核酸酶，则分别用大写罗马数字 I、II、III 等来表示。例如，从流感嗜血菌 d（*Haemophilus influenzae* d）中提取的第三种限制性内切核酸酶被命名为 *Hin*dIII（见下画线处）。

2. 限制性内切核酸酶（专指 II 型酶，下同）的识别序列与切割

绝大多数的限制性内切核酸酶识别 4～6 个核苷酸长度的序列，极少数酶识

别更长的序列，这些序列是旋转对称的（即呈回文结构）（表 15-2）。不同酶的切割位点有所不同：一些酶可以从识别序列的对称中心处切割，产生的 DNA 片段带平齐末端，如 *Hae*III，识别序列是 5′-GGCC-3′，切割位点在 G、C 之间，切割后是平齐末端［图 15-6（a）］；另一些酶在序列对称轴外的两个对称点切开 DNA 的两条链，切割后产生黏性末端，如 *Eco*R Ⅰ 的识别序列是 5′-GAATTC-3′，产生的 DNA 是黏性末端［图 15-6（b）］。

5′-GG↓CC-3′ / 3′-CC↑GG-5′ —*Hae* III→ 5′-GG / 3′-CC + CC-3′ / GG-5′

(a) *Hae* III 酶切位点

5′-G↓AATTC-3′ / 3′-CTTAA↑G-5′ —*Eco*R Ⅰ→ 5′-G / 3′-CTTAA + AATTC-3′ / G-5′

(b) *Eco*R Ⅰ 酶切位点

图 15-6　不同限制性内切核酸酶切割后产生不同的末端

至今已分离到 400 多种限制性内切核酸酶，表 15-2 中列出了一些常用的限制性内切核酸酶。有些酶具有相同的识别序列，称为同裂酶。同裂酶可以具有相同的切割位点，称为同裂同功酶；也可以具有不同的切割位点。有些同裂酶对于切割位点上的甲基化碱基敏感度不同，常常用于 DNA 甲基化作用的研究。还有一些限制性内切核酸酶识别序列不同，但切割后产生相同的黏性末端，称为同尾酶，如 *Bam*HI 和 *Bgl*II。同尾酶产生的相同黏性末端可以彼此互补相连，经常被用于基因克隆实验中。

表 15-2　一些常用的限制性内切核酸酶

限制性内切核酸酶	识别序列和切割位点	限制性内切核酸酶	识别序列和切割位点
*Bam*HI	—G↓GATCC— —CCTAG↑G—	*Bgl*II	—A↓GATCT— —TCTAG↑A—
*Eco*RI	—G↓AATTC— —CTTAA↑G—	*Hin*dIII	—A↓AGCTT— —TTCGA↑A—
Hpa I	—GTT↓AAC— —CAA↑TTG—	*Kpn* I	—G↓GTACC— —CCATG↑G—
Nde I	—CA↓TATG— —GTAT↑AC—	*Nhe* I	—G↓CTAGC— —CGATC↑G—
Not I	—GC↓GGCCGC— —CGCCGG↑CG—	*Pst* I	—CTGCA↓G— —G↑ACGTC—
Sna BI	—TAC↓GTA— —ATG↑CAT—	*Xba* II	—T↓CTAGA— —AGATC↑T—
Xho I	—C↓TCGAG— —GAGCT↑C—		

3. 限制性内切核酸酶的反应条件

所有限制性内切核酸酶的作用都需要镁离子。绝大多数酶作用的最适 pH 在 7.5～8.0 之间，最适温度为 37℃。不同的限制性内切核酸酶酶表达最大活力的条件有所不同，主要体现在对 NaCl 浓度要求差别较大，另外，巯基乙醇和二硫苏糖醇可以稳定酶的活性，经常被使用。现在限制性内切核酸酶基本上都已商品化，按照说明书进行操作即可。有些酶在非标准条件下，识别序列的特异性降低或发生改变，能够切割非特异的识别序列，这种现象称星号活性，在相应限制性内切核酸酶的右上角加（*）表示。酶的星号活性可造成切割位点的大量增加，在实验中应避免。有关限制性内切核酸酶的使用和注意事项在第三节详述。

（二）DNA 连接酶（ligase）

最常用的是 T4DNA 连接酶，是在 T4 噬菌体感染的大肠杆菌中发现和分离的。它可以催化两条 DNA 链的 3′-OH 和 5′-PO_4^{3-} 之间形成磷酸二酯键，从而把两段 DNA 连接起来。T4DNA 连接酶不仅适用于 DNA 之间的黏性末端连接，也可用于平齐末端的连接，只是对于前者的连接效率要比对后者高很多。

T4DNA 连接酶催化的反应需要 Mg^{2+} 和 ATP。最适 pH 为 7.5～7.6。就酶活力而言是在温度 37℃时最高，但考虑到连接反应时 DNA 的稳定性和 DNA 片段之间的末端相互作用，在实验中进行连接时，平齐末端一般采用 20～25℃，黏性末端 12～16℃。

（三）其他主要工具酶

1. 大肠杆菌 DNA 聚合酶Ⅰ（DNA ploymerase Ⅰ）

此酶由一条分子质量 109 000Da 的多肽组成，是一多功能酶。具有三种活性：5′→3′DNA 聚合酶活性、5′→3′外切核酸酶活性和 3′→5′外切核酸酶活性。主要用于缺口平移法制备 DNA 探针，也可用于 3′凸出的黏性末端的平端化和标记。

2. 大肠杆菌 DNA 聚合酶Ⅰ的 Klenow 片段

大肠杆菌 DNA 聚合酶Ⅰ可被枯草杆菌蛋白酶或胰蛋白酶切割成相对分子质量分别为 76 000 和 36 000 的两个片段，其中大片段称 Klenow 片段，它保留了大肠杆菌 DNA 聚合酶Ⅰ的 5′→3′DNA 聚合酶活性和 3′→5′外切核酸酶活性，而小片段保留 5′→3′外切核酸酶活性。Klenow 片段的 5′→3′DNA 聚合酶活性可用于双脱氧 DNA 序列测定、cDNA 链中第二链的合成以及补齐或标记带 3′凹端的

DNA 片段，其 3′→5′外切核酸酶活性也可用于 3′凸出的黏性末端的平端化。也可用于以 M13 单链 DNA 为模板的 oligo 诱导突变实验中突变链的合成。

3．*Taq* DNA 聚合酶

分子质量 95 000Da，具有强的 5′→3′ DNA 聚合酶活性。最适反应温度 75～80℃，即使 95℃处理仍保持活性。主要用于 PCR 反应。

4．反转录酶（reverse transcriptase RTase）

由两条肽链组成，其中一条同时具有 5′→3′DNA 聚合酶活性和 5′→3′RNA 外切核酸酶活性。它可以以单链 RNA 或单链 DNA 为模板，在引物参与下催化合成与模板互补的 DNA 子链。主要用于以 mRNA 为模板的 cDNA 第一链的合成。

5．核酸酶 S1

在适当的酶浓度下，可特异地降解单链 DNA 或单链 RNA，而对双链 DNA、双链 RNA 以及 DNA-RNA 杂交分子没有作用，在高浓度下也可降解双链核酸。常用于 DNA 黏性末端的平齐、打开双链 DNA 中的发夹环、分析 DNA-RNA 的杂交结构（S1 mapping）。

6．核糖核酸酶 H

是一种内切核酸酶，可特异地降解 DNA-RNA 杂合链的 RNA 链。在 cDNA 的第一链合成完成后，常用核糖核酸酶 H 去除 mRNA 模板以进行下步反应。

7．碱性磷酸酶（alkaline phosphatase）

该酶在碱性条件下，可以催化 DNA 或 RNA 的 5′端脱去磷酸基团。在 DNA5′端进行放射性标记时，先通过此酶产生 5′-OH，再通过核苷酸激酶的作用将放射性核苷酸加到 5′端进行标记；此酶还有一个用途，在载体和目的基因的连接反应中去除载体 5′端的磷酸基团，防止载体自连。目前常用的有两种酶：细菌碱性磷酸酶（BAP）和牛小肠碱性磷酸酶（CIP）。

8．末端脱氧核苷酸转移酶（terminal deoxynuleotidyl transferase）

这种酶可以将脱氧核苷酸反复地加在单链和双链 DNA 的 3′-OH 端。如果是单链 DNA 或 3′-OH 端凸出的双链 DNA，反应在 Mg^{2+} 存在的条件下进行；如果是平齐末端或 3′-OH 端凹进的双链 DNA，则需要 Co^{2+}。末端脱氧核苷酸转移酶主要用于在载体和外源 DNA 片段上加上可以互补的同聚体尾部（如在载体两端加上 AAAAA，在外源片断上加上 TTTTT），以便两者能在体外进行重组。另外，此酶也常用于 DNA 片段的 3′端标记。

9．T4 多核苷酸激酶

具有两种活性。可以将 ATP 的 γ-磷酸转移到 DNA 或 RNA 的 5′端羟基，这种正向反应活性是高效率的，常用于 DNA、RNA 5′端标记和在 oligo DNA（引物）5′端加上磷酸基团。另外一个活性是交换活性，催化 5′端磷酸的交换，即在 ADP 过量存在的条件下，可以将 5′端的磷酸基团转移到 ADP 上，这种活性很低，但可以用于 DNA、RNA 5′端的去磷酸化，以便加上标记的磷酸。

第二节　目的基因制备

人们将那些已被分离或欲分离、扩增、表达、改造的基因或 DNA 片段称为目的基因。显然“目的基因”不仅仅指那些能编码蛋白质的 DNA 片段，它是一个包含了诸如启动子、终止子、增强子等调控序列以及非编码区域在内的一个广泛的定义。人类细胞的基因数目可能高到几万个，即使是简单的大肠杆菌也有几千个基因，如何从如此庞大的基因组中获取我们所需要的目的基因是重组 DNA 技术的第一个难题。随着分子生物学及其相关技术的发展，人们已经用多种多样的方法分离目的基因，主要有直接分离法、构建基因组文库或 cDNA 文库分离法、PCR 法和化学合成法等。

一、直接分离法

1. 基因分离的物理化学法

这是重组 DNA 技术发展初期的方法，基本原理是 DNA 分子的 GC 碱基配对中存在 3 个氢键、AT 间只有两个氢键。如果不同基因片段间的碱基组成差异较大，其理化性质如浮力密度、解链温度等就有明显不同，采用相应的办法即可分离到目的基因。这种方法可以分离到某些生物的 rDNA 基因，但目前已很少应用此方法分离目的基因。

2. 限制性内切核酸酶酶切分离

适用于从简单基因组中分离含目的基因的 DNA 片段。质粒和病毒等的 DNA 分子只有几千到几十万个碱基，目的基因较少，获得目的基因的方法也比较简单，但是，由于受限制性内切核酸酶位点要求的限制，难以获得较为理想的片段。

3. 利用反转录法直接从特定 mRNA 上分离基因

主要用于转录产物 mRNA 以分离目的基因。以目的基因的 mRNA 为模板，在反转录酶的作用下合成互补的 cDNA，然后在 DNA 聚合酶的作用下合成双链

DNA，与适当的载体重组后转入宿主细胞扩增，即可获得目的基因的 cDNA 克隆。对于一些低丰度 mRNA，直接用细胞的总 mRNA 为模板往往难以得到cDNA克隆，现在发展出的双抗体免疫分离目的基因技术解决了这个问题，其主要原理是利用已纯化的蛋白、蛋白抗体以及蛋白抗体的第二抗体，通过蔗糖密度离心等方法将含有特定 mRNA 的多聚核糖体从总多聚核糖体中分离出来，利用反转录法获得目的基因。

二、构建基因组文库或 cDNA 基因文库分离法

对于高等真核生物而言，基因组 DNA 十分庞大，且基因组成及结构复杂：除了编码序列，还有调控序列、间隔序列、重复序列等，绝大多数基因难以直接分离。一个可行的方法是对所有的基因进行克隆和扩增，也就是建立某种生物的基因组文库或 cDNA 文库，然后根据不同的方法将所需的克隆分离出来，最终分离获得目的基因。从 cDNA 文库中获得基因的编码区，可以避免真核基因过大或者含有多个内含子的麻烦。

1. 基因组文库

基因组文库是指将某种生物基因组上所有的遗传信息通过克隆储存在一个受体菌的群体中。若受体群体中只储存了部分的遗传信息，则称部分基因组文库。染色体上的 DNA 是生物遗传信息的主要携带者，将其比较均匀地“剪断”是构建基因组文库的第一步，在早期人们用超声波等物理方法或限制性内切核酸酶等生化方法将染色体 DNA 打断，通过与适当的载体相连得到数目众多的重组体，再用探针杂交等方法将目的基因“钓”出，这就是所谓的“鸟枪法”，这种方法在很大程度上是依赖于运气，因为目的基因，尤其是单拷贝基因，在整个基因组中的比例是非常低的。在以后的发展中，这项技术不断得到完善，包括特异性更适合的限制性内切核酸酶、能容纳更大片段的载体（如噬菌体载体、cosmid 载体和 YAC 载体）、更先进的筛选鉴定方法等，从而使目的基因的分离主要是依赖于技术手段而非运气，也使构建基因组文库成为现代重组 DNA 技术中的一个重要内容。

2. cDNA 基因文库

将某种生物基因组转录的部分或全部 mRNA 经反转录法产生的各种 cDNA 和适当的载体相连，储存于一种受体菌群体中，称为构建 cDNA 基因文库。在细胞中，表达基因一般仅占基因总数的 15%左右，所以从 mRNA 出发分离目的基因极大地缩小了搜索目的基因的范围。与基因组文库相比，cDNA 基因文库有其自身的特点：①cDNA 序列只反应基因组的编码序列，文库的复杂性较低。②生

物在不同的发育阶段或特定的组织中，表达的基因有所不同，因而构建的 cDNA 文库也会有差别。③从 cDNA 文库筛选的目的基因可直接用于表达。

三、PCR 法分离目的基因

如果知道目的基因的全序列或其两侧序列，可以通过合成一对与模板 DNA 互补的引物，采用 PCR 方法即可有效地分离目的基因。与传统的 DNA 克隆方法相比 PCR 方法显得十分简捷。现代分子生物学发展到今天，人类基因组全序列已经测出，并且越来越多的模式生物的基因组全序列正在被测出，PCR 法已经成为分离目的基因的主要手段。

四、目的基因的化学合成法

是指根据已知的或设计的蛋白质氨基酸的顺序，推断出目的基因的碱基顺序，通过化学合成的方法合成目的基因。化学合成法适用于小 DNA 分子质量的基因，也可以进一步通过酶学方法再连接出较大的基因。随着结构生物学和生物信息学的快速发展，DNA 的化学合成法已经成为分子生物学实验室的一种非常重要的常规技术手段。

第三节　目的基因与载体的体外重组

顾名思义，目的基因与载体的体外重组是重组 DNA 技术中的核心步骤。通常所说的体外重组指 DNA 片段之间的体外连接，即由 DNA 连接酶催化两个双链 DNA 片段间的 5′端磷酸基因和 3′端羟基形成磷酸二酯键的过程。目前常用的 DNA 连接酶有两类，一类来源于大肠杆菌，另一类来源于 T4 噬菌体。其中以 T4 连接酶最常用，它既可以催化带有互补黏末端的 DNA 片段的连接，也可以用于平齐末端的 DNA 片段的连接。DNA 体外重组主要由这两种连接方式进行。体外重组技术发展到现在，已经与限制性内切核酸的酶切作用密不可分了。无论是 DNA 片段间的互补黏性末端，抑或是平端末端，绝大多数由限制性内切核酸酶酶切而得（当然，还包括一些其他辅助手段）。因此，我们将目的基因与载体的剪切纳入此节叙述。

一、目的基因与载体的剪切

目的基因与载体的剪切一般通过限制性内切核酸酶进行切割。即使获得的目的基因片段没有合适的酶切位点，也可以通过 PCR 等常规方法在其两端加上。

一个普遍性的原则是采用相同的限制性内切核酸酶（或同尾酶）分别切割目的基因和载体，使两者的末端能够互补，便于下一步的连接反应。目的基因和载体的剪切主要有以下几种方式：

（一）限制性内切核酸酶切割后形成黏性末端

由限制性内切核酸酶切割形成的黏性末端有两种：5′凸出末端（如 *Bam*HⅠ切割后）和 3′凸出末端（如 *Pst* Ⅰ切割后）。两种黏性末端都可以使用，但前者的连接效率更高。限制性内切核酸酶有几百种，其选择方法很简单：先看载体上有哪些合适的单一酶切位点，再看这些酶切位点有哪些不包含在目的基因中，即可被选用。当然，有时构建一些特殊载体如表达型载体时还要考虑避免移框、符合表达型载体的启动子要求等问题，这一般在载体的使用说明中有详细叙述。

在第一节中已简要提及限制性内切核酸酶的一些使用要求，以下具体谈谈影响限制性内切核酸酶的几个因素。

1. DNA 样品的质量

一般而言，经过碱法提取的 DNA 样品，大多数限制性内切核酸酶都可以直接使用，并且有较好的酶切效果。但 DNA 样品的某些污染物，如苯酚、氯仿、乙醇、EDTA，SDS、蛋白质以及高浓度的盐离子等，有可能抑制限制性内切核酸酶的活性，因此制备 DNA 样品时首先是要求操作规范。少数限制性内切核酸酶，即使用规范操作提取的 DNA 样品其酶切活性也很低，通常采用增大酶用量、延长反应时间、扩大反应体积或者提高 DNA 纯度等方法来获得满意的切割结果。

有些 DNA 样品，尤其是采用碱变性法制备的，会含有少量的 DNase 的污染。一般 DNA 储存液中含有 EDTA，螯合了 DNase 活性所需的 Mg^{2+}，因此 DNA 是稳定的。但随着限制性内切核酸酶缓冲液中 Mg^{2+} 的加入，有时 DNA 会被 DNase 很快的降解。一般采用提高样品中 DNA 的浓度来避免这种情况。

2. DNA 的甲基化

大肠杆菌有两种甲基化酶，dam 甲基化酶和 dcm 甲基化酶。识别序列中核苷酸的甲基化作用，会强烈的影响大多数限制性内切核酸酶的活性。为避免产生这样的问题，重组 DNA 技术中一般使用失去了甲基化酶的大肠杆菌菌株制备 DNA。

3. 酶切反应体积、酶的用量和反应时间

商品化的酶一般有 50% 的甘油和少量 EDTA。如果反应体积太小，甘油和

EDTA 都会影响到酶的活性。一般切割 1μgDNA 的体积在 20～30μl 为宜，酶用量一般为 1μgDNA 5 单位酶时足够了（注意切割 1μgDNA 的不同限制性内切核酸酶用量要求是不同的)。由于通常酶的用量大于所需量，故反应时间不能过长，一般不超过 2h，以免其他杂酶影响酶切效果。有时酶切的 DNA 较多，体积较大，酶的用量可减少到 1μgDNA 2～3 单位酶，同时把反应时间延长至 3h。

4. 限制性内切核酸酶的缓冲液

第一节已经提及限制性内切核酸酶的反应条件，依据反应条件建立的标准缓冲液的组分包括氯化镁、氯化钠或氯化钾、Tris-HCl、β-巯基乙醇或二硫苏糖醇(DTT)。对于不同的酶，变化最大的是氯化钠或氯化钾的浓度。使用单酶进行酶切时，商业化的酶制剂已经很好地解决了这个问题，不用多作考虑。使用两个或两个以上的酶进行酶切时，可以按产品说明寻找它们的共用缓冲液，每种酶的酶活都达到 50%以上即可。如果没有合适的共用缓冲液，可以考虑先使用低盐浓度缓冲液的酶，再使用高浓度缓冲液的内切核酸酶进行酶切，也可以在一种酶酶切之后，将 DNA 重新提纯、溶解，再用另一种酶酶切。

（二）形成带平末端的 DNA 片段

DNA 片段之间平末端的连接效率不到 5′凸出黏性末端的 1/50。一般在特殊情况下才使用平端连接，如找不到合适的黏性末端限制性内切核酸酶切割目的基因和载体，或是为下一步某些特殊的修饰而使用，有时也是为了基因表达中形成正确的框架连接。

1. 用限制性内切核酸酶获得平末端

一些限制性内切核酸酶，如 *Sna*BI、*Hin*cII、*Hae*II 等可以剪切形成平末端。平端内切核酸酶的反应条件与前述的黏性末端内切核酸酶相同。

2. 由黏性末端改造成平末端

通常有两种方法，补平法和切平法。补平法是用 DNA 聚合酶Ⅰ（Klenow Ⅰ)，以限制性内切核酸酶切割的 5′凸出末端的单链部分为模板，催化 DNA 链形成平末端。而切平法则是用 S 核酸酶或绿豆核酸酶对 5′凸出末端和 3′凸出末端的单链部分进行切割，使 DNA 片段形成平末端，这两种酶要求的切割条件都比较严格，S 核酸酶常常会切割过头而形成“切入”现象，而绿豆核酸酶常常会切割不完全，遗留 1～2 个核苷酸。因此由黏末端改造成平末端多采用补平法。

（三）从平末端改造成黏末端

通常，连接平末端DNA分子的方法除了直接利用T4连接酶进行连接，或是通过PCR方法加上酶切点之外，还可以在平齐末端DNA分子的末端加上一些核苷酸，改造成黏性末端，利用T4连接酶进行连接。常用的方法有同聚物加尾法、接头法和衔接子法。

1. 同聚物加尾法

利用末端脱氧核苷酸转移酶，分别在载体和目的基因的两端分别加上互补的单核苷酸同聚物，如在载体的两端加上poly（dG），在目的基因的两端加上poly（dC），使DNA分子之间具有能够互补的黏性末端便于连接。

2. 接头法

接头是指一类人工合成的双链寡核苷酸片段，它一头具有可与载体连接的某种限制酶的黏性末端，另一头具有平末端。通过平端连接将接头接在目的基因的末端，为避免平端连接时形成接头二聚体，需要将黏性末端的5′端磷酸基团在平端连接前先去除，在平端连接后再用T4多核苷酸激酶加上，进一步进行与载体的连接。

3. 衔接子法

衔接子也是一类人工合成的双链寡核苷酸链。与接头不同的是它两头都是平齐末端，且常常含有两个或两个以上的酶切位点，因此它比接头要长一些。在与目的基因进行平端连接后，可以用适当的限制性内切核酸酶酶切目的基因和载体，使两者产生能互补的黏性末端。

二、目的基因和载体的体外连接

大肠杆菌连接酶只能催化DNA片段间互补黏性末端的连接，不能催化平齐末端的连接，其催化连接反应这一耗能过程的能量来自于氧化型烟酰胺腺嘌呤二核苷酸（NAD^+），因此反应系统中需要添加NAD^+作为辅助因子。T4连接酶既能催化黏性末端之间的连接，也能催化平末端之间的连接，因而比前者更为常用，其催化反应的能量来自于ATP。

连接酶催化连接反应时的最适温度是37℃，但在这个温度下黏性末端之间的氢键结合不稳定。综合考虑酶催化效率和黏性末端结合速率，一般认为实际反应的温度在4～15℃比较合适，反应时间在2h以上甚至过夜。T4连接酶催化平

端连接时，效率比黏性末端要低得多，因此反应温度相应地提高到 22～30℃。

体外连接主要以黏性末端连接和平齐末端连接两种方式为主，根据具体策略的不同，主要有以下几种连接方法。

1. 单一限制性内切核酸酶酶切产生的同源黏性末端之间的连接

目的基因和载体分别用同一种限制性内切核酸酶酶切后，它们两端都带有相同的、彼此之间能够互补的黏性末端（当然也可以用同尾酶酶切得到）。为防止载体自身连接，线性化的载体在连接前需用碱性磷酸酶去除 5′的磷酸基团，在进行连接反应时，连接酶的用量要比不脱磷酸基团时大得多，同时，载体与目的基因的摩尔数量比例一般控制在 1∶5 左右，目的片段再多的话，容易造成多拷贝插入，反而不利。

采用这种连接方法，目的基因是随机连接上载体的，因此会有两种插入方向。对于表达型重组子必须鉴定正确的连接方向，鉴定方法很简单，一般采用目的基因上的一个单酶切位点，辅以载体上的一个单酶切位点，进行限制性酶切图谱分析，根据切出片段的大小即可鉴定方向。同时，这种连接常常会造成目的基因的多拷贝插入，也需要用限制酶酶切图谱进行鉴定。如果电泳中载体的带和插入片段的带所含的 DNA 含量不合比例，就可能是多拷贝插入。

2. 定向克隆

定向克隆是指将目的基因定向的插入到载体中。它要求线性化载体两端（目的基因）的两个末端不能形成互补，只能与目的基因（载体）相应的末端相连接。达到这一要求的 DNA 分子末端有两种形式：①线性化载体（目的基因）两端为非互补的两个黏性末端；②载体（目的基因）两端一侧为黏性末端，另一侧为平末端。对于第一种形式而言，可以通过两种不同的非同尾限制性内切核酸酶分别对载体和目的基因酶切而得，这种连接形式的定向克隆连接效率非常高，是重组连接中最简捷，最常用的方式。至于第二种形式，可以由两种途径得到：用一个产生黏性末端的限制性内切核酸酶和一个产生平齐末端的限制性内切核酸酶对载体和目的基因进行酶切；或者先用一个产生黏性末端的限制性内切核酸酶酶切，用核酸酶切平或用 DNA 聚合酶补齐后，再用另一产生黏性末端的限制性内切核酸酶进行酶切而得。

定向克隆有很多优点：可以防止载体的自身环化、阳性重组子的得率高、对于表达型重组子可以保证目的基因的正确插入。

3. 平端连接

DNA 片段间的平端连接也是重组 DNA 技术中的常用技术。因为很多限制性内切核酸酶酶切可产生平末端、PCR 和机械剪切也可产生平末端，此外还可以

用 DNA 聚合酶补平或核酸酶切平黏性末端，这些不同来源的平末端之间都可以进行连接，这给不同 DNA 分子间的连接带来了极大的方便。

平端连接时要加大 T4 连接酶的用量，一般要达到黏末端连接的 2～5 倍；为防止载体自连一般也要进行去磷酸化处理；载体与目的基因的摩尔数量比在 1∶10左右；并且常加 8%的聚乙二醇以便提高连接效率。与单一限制性内切核酸酶酶切产生的同源黏性末端之间的连接相类似，平端连接中目的基因的插入也是双向的，且可能形成多拷贝，重组子需要进行限制性内切核酸酶酶切鉴定。

第四节　重组子导入细胞技术

体外重组的 DNA 分子，必须进入到合适的宿主细胞（host cell）才能进行扩增和表达。在基因克隆技术中，依据重组载体和宿主细胞的不同，运用不同的名词描述进入过程。质粒 DNA 或以质粒为载体构建的重组子导入细菌细胞，并得以稳定维持或表达的过程称为转化（transformation）。通过病毒导入宿主细胞并建立稳定状态的过程称为转染（transfection），如果病毒是噬菌体，即以噬菌体为媒介，将外源 DNA 导入细胞的过程，又常常被称为转导（transduction）。实际上三者没有本质的区别。

随着基因工程技术的发展，从低等的原核细胞，到简单的真核细胞，进一步到结构复杂的高等动、植物细胞都可以作为重组 DNA 的宿主细胞。至今为止，用作受体菌的原核生物主要有大肠杆菌、枯草杆菌、链霉菌和蓝细菌等，真核生物中也发展了酵母细胞、植物细胞、哺乳动物细胞和昆虫细胞等作为宿主细胞。一般情况下，宿主细胞的选择有如下几个原则：①便于重组 DNA 分子的导入；②没有特异的限制性内切核酸酶系统降解重组 DNA；③导入的重组 DNA 具有较高的遗传稳定性，不易发生变异；④重组 DNA 扩增过程中不对其进行修饰，或尽量接近于目的基因来源的修饰；⑤宿主细胞在遗传密码的应用上没有明显的偏好性；⑥菌株或细胞株的遗传稳定性高，易于扩大培养或发酵生长；⑦安全性高，无致病性。

现已发展了多种方法将重组 DNA 分子导入到宿主细胞中，如冷 $CaCl_2$ 法、类脂法、磷酸钙共沉淀法、电击法、噬菌体介导法等。宿主细胞不同，选用的方法有所不同，下面逐一进行简要介绍。

一、重组 DNA 分子转化原核生物细胞（大肠杆菌）

用原核生物为宿主细胞进行的基因工程操作中，以大肠杆菌的应用最为广泛。重组 DNA 分子转化大肠杆菌细胞主要采用冷 $CaCl_2$ 法和电击法。

1. 冷 $CaCl_2$ 法

转化是指细菌由于吸收了外源 DNA 而发生可遗传表型变化的过程，是自然存在的过程。转化发生时，要求细胞处于能吸收外源 DNA 的生理状态，即感受态（competence）状态。冷 $CaCl_2$ 法利用上述原理，在低温（0℃）、低 Ca^{2+} 强度的条件下，人为地制备感受态细胞，随后加入重组 DNA 分子，它能和 Ca^{2+} 形成抗 DNA 酶的复合物黏附于细胞表面，通过后续的热激处理等步骤而进入细胞。

冷 $CaCl_2$ 法的转化效率较高，1μg 的环状 DNA 分子可以获得 10^5～10^6 个转化子，足以满足一般条件下 T4 连接酶所产生重组子的转化。如果加上 Mg^{2+}、DMSO 等物质的联合处理，可以使冷 $CaCl_2$ 法的转化效率提高 100～1000 倍，用于某些连接效率很低或有特殊要求（如文库构建中）的重组 DNA 分子的转化。

2. 电击法

电击法最早用于将外源 DNA 导入真核细胞，后来也成功地应用于大肠杆菌的转化。电击法不需要预先制备感受态细胞，其原理是利用高压电脉冲作用，在大肠杆菌的细胞膜上进行电穿孔，形成通道，从而促进了外源 DNA 的进入。电击法的转化效率可达 10^9～10^{10}转化子/1μg 环状 DNA。电击法常用于转化分子质量大于 15kb 的重组 DNA 的转化。

除了冷 $CaCl_2$ 法和电击法之外，还有三亲本杂交（tri-parental mating）结合转化法用于重组 DNA 分子转化细菌，它主要适用于那些难以用冷 $CaCl_2$ 法和电击法直接进行转化的受体细菌。

二、重组 DNA 分子导入哺乳动物细胞

外源 DNA 是很难进入哺乳动物细胞的，这在一定程度上影响了哺乳动物基因工程技术的发展。不过，现在已经建立了多种技术能有效地将外源 DNA 导入哺乳动物细胞。

1. 磷酸钙沉淀法

其原理是哺乳动物细胞能捕获黏附在细胞表面的 DNA-磷酸钙沉淀物，使 DNA 转入细胞。磷酸钙沉淀法曾被广泛地应用于 DNA 转染哺乳动物细胞（以及昆虫细胞）。一般情况下转染的细胞可占总数的 10%。磷酸钙沉淀法的操作步骤比较繁琐，试剂要求很严格，尤其是 Hepes 缓冲液和钙盐溶液的 pH 要求准确到 7.05。磷酸钙沉淀法成本低廉，到目前为止仍是比较常用的转染方法之一。

2. 电击法

电击法转染哺乳动物细胞的原理与电击法转化大肠杆菌的相同，不同的是前

者的转染效率要远远低于后者的转化效率，一般在每 10^5 个存活细胞中，可得到 20～30 个转化子，并且转染效率与宿主细胞类型密切相关。由于电击时的介质溶液电阻越大越好（低离子强度溶液），但又不能使细胞在低渗条件下过度膨胀破裂，因此一般情况下是选择合适的电压、介质离子强度以及脉冲时间，使电击后有 50%～60%的细胞死亡，才能获得满意的结果。电击法转染真核细胞的操作过程增加了细胞污染其他杂菌的机会，因此并不常用，不如脂质体法便利和安全。

3. 脂质体法

脂质体是一类人工合成的脂类物质颗粒，可以把外源 DNA 包裹在其中。与裸露的 DNA 分子相比，它与真核细胞细胞膜的脂质双分子层有更好的亲和性，因而脂质体法的转染效率可达磷酸钙沉淀法的一倍以上。由于脂质体无毒、无免疫原性，不仅可以用于体外的细胞转染，而且还可以在动物体内将基因转入靶细胞、靶组织、靶器官位点，实现瞬间表达或稳定表达，因而它具有极大的应用前景。目前，脂质体已经商品化生产，如美国 BRL 公司的 lipofectin 产品。利用脂质体转染哺乳动物细胞已经成为实验室的常规技术。

4. 显微注射转基因技术

应用显微注射器可以把重组 DNA 直接注入哺乳动物细胞内。用这种方法转基因的效率很高，获得稳定转化子的数量主要取决于重组 DNA 的性质。

除了上述四种方法，还可以采用 DEAE-葡聚糖转染法和聚阳离子-DMSO 转染法将外源 DNA 导入哺乳动物细胞，它们的原理与磷酸钙沉淀法类似，葡聚糖和聚阳离子的使用都是为了促进细胞表面对 DNA 的吸附力，DMSO 的使用则是为了增加细胞膜的通透性。此外，病毒颗粒的感染也可实现外源 DNA 的导入。

三、重组 λ 噬菌体 DNA 分子转导大肠杆菌

λ 噬菌体载体主要用于较大的基因片段的克隆。λ 噬菌体本身的基因组就很大，加上外源片段后其大小可达 50kb 左右，这么大的 DNA 分子通过转化进入细胞的可能性非常小。现在运用重组 λ 噬菌体 DNA 分子转导大肠杆菌的技术已经发展成熟，其要点是将重组 λ 噬菌体 DNA 进行体外包装，形成具有感染能力的病毒颗粒，通过感染细胞将外源 DNA 带入细胞内。

四、重组 DNA 分子导入植物细胞

植物细胞有细胞壁的结构，因此重组 DNA 导入的方法与动物细胞有所不

同。

目前在植物细胞中最常用的方法是农杆菌介导的 Ti 质粒载体转化法。农杆菌是一类革兰氏阴性菌，能够通过伤口感染双子叶植物和裸子植物，并将其 Ti 质粒上的 T-DNA（可以包含目的基因片段）转移至细胞内部，与染色体 DNA 整合，从而得到稳定维持或表达。农杆菌介导的 Ti 质粒载体转化法已经成熟，迄今为止约有 80％的转基因植物是通过这种方法得到的。

但是农杆菌对于绝大多数的单子叶植物没有感染力，为了在这类宿主细胞中实现外源基因的导入，至今已发展了电击法、基因枪法、脂质体法、显微注射法、多聚物介导法等方法。这些方法多是参照哺乳动物细胞而来，只不过在多数情形下需要去除植物细胞的细胞壁，以原生质体的形式完成重组 DNA 的导入。

第五节　重组子的筛选与鉴定

在重组 DNA 分子导入到宿主细胞的过程中，一般只有少数重组子能进入宿主细胞并且稳定存在。以质粒转化大肠杆菌为例，转化效率在一般情况下只能达到 10^{-5}～10^{-6}，这意味着如果有 10^7 个宿主受体细胞，只有 10～100 个细胞被转化，这些被转化的细胞与其他非转化细胞是混杂在一起的，并且其中还包括自连的载体或者一个载体与多拷贝目的基因片段的连接等非目标重组子的被转化细胞。因此，筛选与鉴定含有阳性重组子的宿主细胞是重组 DNA 技术中一个重要的环节。

不同克隆载体、宿主细胞类型以及不同的重组 DNA 分子导入细胞方法，其重组子的筛选与鉴定也不尽相同，一般主要包括以下几方面。

一、根据重组载体的选择性标记进行筛选

目前经常使用的载体都携带有选择性遗传标记，或具有一种可供选择的遗传学特征，这样重组载体导入宿主细胞后，可以使宿主细胞呈现出特殊的表型或遗传学特征。据此可对重组子或转化子进行初步的筛选。

1. 抗药性筛选

载体携带的选择性标记中最常见是抗药性标记，如抗氨苄青霉素基因（amp^r）、抗四环素基因（tet^r）、抗卡那霉素基因（kan^r）等。

如果外源基因片段插入载体的位点在抗药性基因之外，不导致抗药性基因的失活，仍能编码抗药性基因产物。含有这种重组子的转化细胞可以在含有相应抗生素的琼脂平板上生长成菌落，那些未能接受载体 DNA 的细胞因不能生长而被筛除，但是除了阳性重组子之外，自身环化的载体、未酶切完全的载体以及非目

的基因插入载体形成的重组子均能导入宿主细胞并进行抗药性生长形成菌落，这些菌落需要用进一步的鉴定进行排除。

如果外源基因片段是插入在载体的抗药性基因中间，导致抗药性基因失活，这个抗药性标志就会消失。例如质粒 pBR322 含有 amp^r 和 tet^r 两个抗药基因，若将目的序列插入 ter^r 基因序列中，转化大肠杆菌，将细菌涂布含氨苄青霉素或四环素的琼脂平板上，未接受质粒 DNA 的细胞都不能生长；在含氨苄青霉素和四环素的琼脂平板中都能生长的细菌是含有空质粒 pBR322 的（未能插入目的基因）；在含氨苄青霉素的琼脂平板中能生长、而在含四环素的琼脂平板中不能生长的细菌就很可能是含有目的基因的重组质粒。

2. 显色互补筛选法

最常用的是 β-半乳糖苷酶系统。一个完整的乳糖操纵子包含有调控序列、β-半乳糖苷酶基因（*lacZ*）、透性酶基因（*lacY*）和乙酰转移酶基因（*lacA*），β-半乳糖苷酶系统是利用其中的 β-半乳糖苷酶的活性特征：当培养基中含有诱导物异丙基-β-D-硫代半乳糖苷（IPTG）和生色底物 X-gal 时，含有调控序列和完整 β-半乳糖苷酶基因（*lacZ*）的细菌或者噬菌体会因蓝色沉淀物的产生而形成蓝色菌落或蓝色噬菌斑。根据这个原理，将此操纵子的调控序列和 β-半乳糖苷酶基因（*lacZ*）分成两部分，载体或噬菌体上含有调控序列以及 β-半乳糖苷酶基因（*lacZ*）N 端 146 个氨基酸（α-肽）的编码序列，并在这个编码序列中插入了一个多克隆位点区，它不破坏操纵子的阅读框，在产生蛋白时只是在 β-半乳糖苷酶的氨基端增加几个氨基酸而不影响功能；操纵子的另一部分包含于宿主细胞的基因组中，质粒和宿主细胞的编码产物可以进行 α-互补而形成有酶学活性的蛋白质，在 IPTG 和 X-gal 存在的条件下可以形成蓝色菌落或者蓝色噬菌斑。但如果外源目的基因插入到质粒的多克隆位点以后，导致了 *lacZ* 的失活，不能形成 α-互补，在含有 IPTG 和 X-gal 的平板上形成的是白色菌落或者无色噬菌斑。这样经目测即可对阳性重组子进行初步筛选。

目前使用的很多载体带有 β-半乳糖苷酶基因（*lacZ*）部分缺失的片段，采用 β-半乳糖苷酶系统的显色进行筛选在近年来应用甚为广泛。

3. 插入表达筛选法

一些载体在筛选标记基因的上游连接一段负调控序列，当外源片段插入负调控序列使其失活时，筛选标记基因才能得以表达。例如 pTR262 质粒，它由 pBR322 衍生而来，其标记基因 tet^r 上游存在 *cI* 的负调控序列，*cI* 基因表达的阻遏蛋白可以抑制 tet^r 基因的表达。当外源片段插入 *cI* 中的 *Hin*dIII 位点，可以造成 *cI* 基因的失活而解除了对 tet^r 基因的抑制，从而使阳性重组子呈现 tet^r 表型。

二、PCR 法

PCR 技术的出现给克隆的筛选增加了一个新手段。如果已知目的序列的长度和两端的序列，则可以设计一对引物，以转化细胞的质粒 DNA 为模板或以重组后的细胞基因组 DNA 为模板进行 PCR 扩增，若能得到预期长度的 PCR 产物，则该转化细胞就可能含有目的序列。

三、限制性内切核酸酶酶切分析法

利用限制性内切核酸酶酶切分析法可以进一步筛选鉴定重组子，而且能判断外源 DNA 片段的插入方向以及大小。其基本方法是快速提取转化细菌中的质粒 DNA 后，选用合适的限制性内切核酸酶对质粒 DNA 进行单、双酶切，再通过凝胶电泳分析外源基因的插入方向和大小。

四、核酸杂交法

1. 菌落（或噬菌斑）杂交筛选法

利用标记的核酸作为探针与转化细胞中的 DNA 进行分子杂交，可以直接筛选和鉴定阳性重组子。常用的方法是将转化后生长的菌落复印到硝酸纤维素膜上，用碱裂解细菌，菌落释放的 DNA 就吸附在膜上，再用标记的核酸探针进行分子杂交，核酸探针就结合在含有目的序列的菌落 DNA 上而不被洗脱。核酸探针可以用放射性元素标记，结合了放射性核酸探针的菌落可用放射性自显影法显示，也可以用非放射性物质标记，通常是颜色反应显示，这样就可以将阳性重组子挑选出来。本方法可以进行大规模操作，一次可以筛选 10^6 数量级的菌落或噬菌斑，是从基因文库中挑选阳性重组子的首选方法。

2. Southern 印迹杂交法

为了进一步确定外源 DNA 片段插入的正确性，在使用限制性内切核酸酶酶切、凝胶电泳分离 DNA 片段以后，可以将这些片段转移到硝酸纤维素膜或尼龙膜上，用相应的放射性元素标记或非放射性物质标记的 DNA 探针进行分子杂交，鉴定重组子的插入片段是否是目的基因片段。

核酸杂交法中还包括 Northtern 印迹杂交、斑点印迹杂交、狭线印迹杂交等方法，它们的原理都是利用核酸分子杂交。

五、免疫学方法

利用特定抗体与目的基因表达产物特异性结合的相互作用进行筛选。此法不是直接筛选目的基因，而是通过与基因表达产物的相互作用显示含有目的基因的转化细胞，因而要求实验设计要使目的基因进入受体细胞后能够表达出其编码产物。抗体可以用放射性元素（一般用^{125}I）标记，也可用特定的酶如过氧化物酶、碱性磷酸酶等标记。放射性元素标记抗体在免疫反应后用放射自显影显示，酶标抗体则是利用酶可催化特定底物分解呈现的颜色来显示。免疫学方法专一性强、灵敏度高，因为只要有一个拷贝的目的基因在克隆细胞中表达蛋白质，就可以被检测出来，免疫学方法适用于从大量转化细胞中筛选出极少的含目的基因的细胞克隆。

六、核苷酸序列测定

所得到的目的序列或基因的克隆，一般都要用其核酸序列测定来最后鉴定。已知序列的核酸克隆要经序列测定确证所获得的克隆准确无误；未知序列的核酸克隆需要测定序列才能确知其结构、推测其功能，用于进一步的研究。目前，核酸序列测定已是分子克隆中一个常规的鉴定步骤。核酸序列测定的原理和方法在实验教材中有详细的叙述。

七、植物转化细胞和哺乳动物转化细胞的筛选鉴定

在植物转基因和哺乳动物转基因研究中，常常利用载体携带的报告基因来进行筛选鉴定。报告基因是指一些抗生素抗性基因和一些功能研究得很透彻的编码某些酶或特殊产物的基因。在没有选择压力的条件下，可利用报告基因在受体细胞内的表达，从非转化克隆中选择出转化细胞；同时报告基因可以与目的基因形成嵌合基因，从它的表达可以了解目的基因的表达情况以及研究基因的表达调控。在转基因植物研究中，常用的报告基因有新霉素磷酸转移酶基因、潮霉素转移酶基因、氯霉素乙酰转移酶基因以及荧光素酶基因等，在哺乳动物转基因研究中，常用的报告基因有新霉素磷酸转移酶基因、氯霉素乙酰转移酶基因，胸腺核苷激酶基因和二氢叶酸还原酶基因等。

第六节　克隆基因的表达

克隆基因的表达对于理论研究和实际应用有着十分重要的意义。克隆的基因

只有通过表达才能研究基因功能以及表达调控的机理；一些有特定生物活性的蛋白质，在医学、工业和农业上具有应用价值，可以克隆其基因，并使之在宿主细胞中大量表达而进行规模化生产。要使克隆基因在宿主细胞中表达，就要将它放入带有基因表达所需要的各种元件的载体中，即表达载体（expression vector）之中。克隆基因可以放在不同的宿主细胞中表达，按宿主细胞可将表达系统分为原核生物表达系统和真核生物表达系统（表 15-3），对不同的表达系统，需要构建不同的表达载体。克隆基因在不同系统中的成功表达并没有绝对的把握性，这主要取决于对这些系统中基因表达调控规律的认识程度。

表 15-3　目前在基因工程中主要的外源基因表达系统

原核生物表达系统	真核生物表达系统
大肠杆菌表达系统	酵母表达系统
芽孢杆菌表达系统	昆虫细胞表达系统
链霉菌表达系统	哺乳动物细胞表达系统
蓝藻表达系统	植物基因表达系统

一、目的基因在原核生物表达系统中的表达（大肠杆菌表达系统）

利用原核生物作为基因的表达系统具有很多优点：①原核生物多为单细胞异养，生长繁殖快，可通过发酵快速获得大量基因表达产物。②基因组结构简单，遗传背景研究清楚，便于基因操作。③基因表达调控机制和代谢途径研究较为透彻。④大多含有质粒和噬菌体，便于构建表达载体。

目前在原核生物表达系统中表达的外源蛋白绝大多数是在大肠杆菌中实现的，因此我们以大肠杆菌表达系统为例来介绍目的基因在原核生物表达系统中的表达，在其他原核生物表达系统中，表达外源基因的原理与大肠杆菌大致相似，只是某些蛋白适于在非大肠杆菌表达系统中表达。

大肠杆菌是目前应用最广泛的蛋白质表达系统，也是人们表达外源蛋白的首选表达系统。人们用大肠杆菌作为外源基因的表达工具已有 20 多年的经验积累，对其生物特性和遗传背景有了充分了解。大肠杆菌表达外源基因产物的水平高于其他基因表达系统，表达的目的蛋白量一般可达细菌总蛋白量的 5%～30%，有时甚至能高达 80%。重组 DNA 技术发展之初，人们乐观地认为能够按照设计在大肠杆菌中表达任何蛋白，但是事实并非如此，实际上迄今为止只有极少数蛋白质在大肠杆菌中表达成功且进行商品化生产。其根本原因在于人们还未清楚地了解大肠杆菌的一些生理机制，特别是蛋白质折叠方面的机制。这使得很多时候的外源基因表达成为一个摸索条件，穿越“黑箱”的过程。多年的大量探索使人们

得到了一些规律性的认识。

1. 尽量多试验不同的选择载体-宿主表达系统

在大肠杆菌中获得成功表达的（有活性）多是些小的分泌蛋白，当蛋白长度超过 100 个氨基酸残基时，往往表达不稳定或是形成包含体。现在已经发展出多种系列的载体-宿主表达系统用于蛋白表达，尽管宿主系统的遗传背景已经研究得比较清楚，但是还是不能解释一些蛋白可以在某些菌株中高效表达而在其他菌株却不能。因此，实验选择表达外源蛋白的宿主系统往往成为蛋白表达的唯一标准。

2. 含有融合基因的载体的使用

这类载体在外源基因克隆区已经含有一个或一个以上的已知蛋白（称运载蛋白）的可读框，外源基因克隆后，载体表达的产物是一个融合蛋白。运载蛋白对于靶蛋白的表达和纯化极为有利：运载蛋白可以保护靶蛋白免受宿主细胞的蛋白酶的降解，可以改善靶蛋白的溶解性，防止形成不溶性的包含体，如果运载蛋白是一种信号肽，还可以将融合蛋白引导到特定的细胞区域，此外，运载蛋白可以提供特殊的酶功能或是抗原组成的结构域，用于靶蛋白的标记与纯化。目前使用的商品化表达载体，几乎都包含了运载蛋白，如 pGEX 载体系列中的谷胱甘肽-S-转移酶（GST）、pET 质粒系列中的聚组氨酸标签（6-His Tag）、pMAL 质粒系列的麦芽糖结合蛋白（MBP）等。

3. 密码子的使用

遗传密码与氨基酸不是一对一的关系，20 种氨基酸由 61 个密码子编码，除了两种氨基酸由单一密码子编码外，其他 18 种都各有几种密码子编码。有时需要表达的外源基因含有大量的稀有密码子（相对于细菌宿主系统而言），会严重影响到翻译过程，这种情况下需要通过突变或者重新合成基因除去稀有密码子，从而提高表达水平。

4. 培养条件的优化

即使是相同的表达系统，不同的培养基可以导致表达水平的巨大差异。一般情况下，可以使用 LB 培养基建立表达的基本参数，在对表达条件进行优化时，或是在 LB 培养基中表达水平很低甚至没有表达时，就需要试用其他培养基，包括基本盐培养基，如 M9；限定盐培养基，或是丰富培养基，如 2YT、NZCYM、肉汤；及一些诱导培养基。此外，一些培养基添加剂也能提高某一种特定蛋白的表达水平，如 NaCl、蔗糖、抗生素、血红素等。

培养过程中的其他条件对于外源基因的表达影响也比较大，如培养温度、培

养时的转速、诱导时间、培养液的 pH 等，其根本原因在于细菌自身的生长旺盛时，它的表达/加工系统也会过载，往往会形成没有活性的包含体，有时甚至会阻碍外源蛋白的表达。因此，在诱导的培养过程中，往往是采用非最适合细菌生长的条件，如降低培养温度、降低转速、延长诱导时间等，获得更多有活性蛋白质的表达。

二、目的基因在真核生物表达系统中的表达

利用原核系统表达真核基因存在许多缺陷。①没有真核转录后加工的功能，不能进行 mRNA 的剪接，所以只能表达 cDNA 而不能表达真核的基因组基因；②没有真核翻译后加工的功能，表达产生的蛋白质，不能进行糖基化、磷酸化等修饰，难以形成正确的二硫键配对和空间构象折叠，因而产生的蛋白质常常没有足够的生物学活性；③表达的蛋白质经常不溶，会在细菌内聚集成包含体（inclusion body），尤其当表达目的蛋白量超过细菌体总蛋白量 10％时。无生物活性的不溶蛋白，需要经过复性（renaturation）处理，使其重新散开、重新折叠成具有天然蛋白构象和良好生物活性的蛋白质，但是多数蛋白质的复性率通常很低。

要表达真核生物的蛋白质，采用真核表达系统比原核系统优越，用于表达的真核细胞有酵母细胞、昆虫细胞、以及一些哺乳类动物细胞等。真核表达载体至少要含两类序列：①原核质粒的序列，包括在大肠杆菌中起作用的复制起始序列、能用在细菌中筛选克隆的抗药性基因标志等，以便插入真核基因后能先在操作很方便的大肠杆菌系统中筛选获得目的重组 DNA 克隆，并复制繁殖得到足够使用的数量。②在真核宿主细胞中表达重组基因所需要的元件，包括启动子、增强子、转录终止和加 polyA 信号序列、mRNA 剪接信号序列、能在宿主细胞中复制或增殖的序列，能用在宿主细胞中筛选的标志基因，以及供外源基因插入的单一限制性内切核酸酶识别位点等。本节只是简要介绍酵母表达系统、哺乳动物细胞表达系统和昆虫细胞表达系统。

1. 酵母表达系统

酵母是单细胞真核生物，其生长代谢与原核生物如大肠杆菌等相似，但在基因的表达调控方面类似于高等的真核生物，因此酵母表达系统不仅具有与 *E. coli* 表达系统相同的简便、表达量高、能大规模发酵的特点，还兼顾到蛋白质生物活性的需要。目前酵母表达系统是外源基因最理想的真核生物基因表达系统。现已构建了多种酵母基因表达载体，如自主型复制质粒（穿梭质粒：酵母和大肠杆菌）、整合型质粒、着丝粒型质粒、酵母人工染色体等。主要的宿主菌有酿酒酵母、巴斯德毕赤酵母、乳酸克努维酵母等。其中巴斯德毕赤酵母表达系统是近年来发展起来的高效表达系统，成功地表达了多种外源蛋白。

2. 哺乳动物细胞表达系统

哺乳动物细胞也可用于表达外源基因，目前主要应用于生物制药方面，如生产基因工程疫苗和多肽产品。来源于哺乳动物的外源基因在哺乳动物细胞中表达具有优越性：所表达蛋白质的免疫原性和生物学活性往往高于其他表达系统所表达的。哺乳动物表达系统的表达载体包括质粒载体和病毒载体，质粒载体是一类穿梭质粒，如前所述它包括了使质粒能在原核、真核生物细胞中生长以及在真核生物细胞中表达的各种原核、真核元件，至于病毒载体，目前已有多种病毒如SV40病毒、痘病毒、腺病毒、疱疹病毒、反转录病毒等用作真核细胞表达载体。哺乳动物细胞表达系统的宿主细胞主要是一类与正常细胞生理特征相接近的肿瘤细胞，如 CHO-K1 细胞、COS 细胞、鼠骨髓瘤细胞等。

3. 昆虫细胞表达系统

昆虫细胞表达系统是利用昆虫细胞作为表达的宿主细胞。昆虫细胞表达系统表达的真核外源蛋白在抗原性和生物学活性方面与天然蛋白十分相似。昆虫细胞表达系统同样也具有易操作、安全的优点。已有十几种外源真核基因在昆虫细胞中表达成功，是一种极具发展前途的真核表达系统。以往昆虫细胞表达系统的载体主要是重组昆虫杆状病毒，如苜蓿银纹夜蛾多角体病毒（AcNPV），它是利用外源基因取代病毒的多角体基因，从而将外源基因置于强大的病毒多角体蛋白基因启动子的下游，引发外源基因的表达。目前，已发展出商品化的穿梭质粒作为昆虫细胞的表达载体，较重组杆状病毒载体而言，不需要进行空斑纯化这个鉴定重组病毒的繁琐过程，使昆虫细胞表达系统更加简便易行。杆状病毒的基因组较大（130kb），可以插入大片段的外源 DNA（可达 100kb），并且允许克隆多拷贝基因来提高表达效率。此系统有一个缺点，就是外源蛋白的表达发生在重组杆状病毒原液对细胞的急性溶解感染期，蛋白质的生产是一短期行为，且导致了细胞的死亡，需要不停地制备病毒原液和用于感染的新细胞。

第七节　重组 DNA 技术应用

从 20 世纪 70 年代发展起来的重组 DNA 技术现已成为现代生物技术的核心技术，同时它在生产实际应用方面也取得了惊人的成绩。它的应用范围非常广泛，涉及工业、农业、医学、环境、能源等诸多领域，可以说重组 DNA 技术已进入到人类生活的各个领域。目前，从大的领域分类来谈重组 DNA 技术应用显然为时过早，因为重组 DNA 技术处于一个加速发展的过程之中，现在的重组 DNA 技术应用仅仅只是一个起步。因此，我们仅就重组 DNA 技术应用相对而言较为成熟的医学、农业、环保等领域来阐述重组 DNA 技术的应用。

一、重组 DNA 技术在医学上的应用

作为分子生物学发展的重要组成部分，DNA 重组及基因工程技术给生命科学带来了革命性变化，促进着生命科学各学科研究和应用的进步，对推动医学各领域的发展同样起着重要的作用。

1. 对人类遗传信息的研究

遗传信息决定生物的形态和特征，是生物生存之本。估计人类的基因组 DNA 约有 4×10^9bp，含有约 3 万个基因，但至今人类对自己赖以生存繁衍的这个庞大的遗传信息库还知之甚少，目前已经知道的人基因只占估计数的百分之十，已搞清楚其表达调控者更寥寥无几，对占基因组 80%～90% 不是蛋白质编码序列的认识更少，因而实际上我们现在对自己生存的基础和实质只有很表面的肤浅认识，设想如果人类掌握了自身全部遗传信息的结构、功能、表达和调控，无疑将能够深刻认识人的生长、发育、生存、繁衍的整个生老病死历程，将能对疾病的诊断、治疗和预防提出极有效的措施，将能真正掌握自己生存和发展的命运。

DNA 重组技术的出现和发展，就使人们有可能去深入探索这个重大的课题。1985 年提出的人基因组研究计划（Human Genome Project）很快得到世界科学界的响应，这个研究计划的目标是要阐明人类遗传信息的组成和表达，是迄今全球性生物学、医学领域最引人注目的巨大研究工程。DNA 重组是完成这个任务的主要手段，其中包括大片段 DNA 克隆、DNA 的大尺度分析、全自动 DNA 序列测定、基因组信息数据库的建立等新思维和新技术的不断出现和发展，再加上大规模引入其他领域先进的科学技术，原预定 21 世纪头 10 年绘制出完整的人类染色体基因定位图、测定出人类基因组全部 DNA 序列，业已提前完成。当然在此基础上要搞清楚全部人类基因的功能、各基因间的关系，基因表达调控、人类遗传信息的多样性等还要经历更长期和更艰苦的努力。但 DNA 重组技术促进了分子生物学迅速发展，给人类探索自身生命的奥秘展示了光明的前景。

生命关键的基础在于蛋白质与蛋白质、蛋白质与核酸的相互作用，生物大分子的结构与功能的联系正是生命的本质所在。凭借基因工程人们可以克隆获得天然的或任意设计的核酸序列，可以大量获得过去难以得到的生物体内极微量的活性蛋白质、可以设计获得任意定点突变（site-directed mutagenesis）的基因和蛋白质，这就为研究蛋白质与核酸的结构与功能、揭露生命的本质提供了最基本的手段。

2. 基因工程药物和疫苗

自从 20 世纪 80 年代初第一种基因工程产品——人胰岛素投放市场以来，以基因工程药物为主导的基因工程应用产业已成为全球发展最快的产业之一。利用

基因工程技术生产有应用价值的药物是当今医药发展的一个重要方向，利用基因工程技术生产药物有两个不同的途径：一是利用基因工程技术改造传统的制药工业，例如用DNA重组技术改造制药所需要的菌种或创建的菌种，提高抗菌素、维生素、氨基酸等的产量；二是用克隆的基因表达生产有用的肽类和蛋白质药物或疫苗。目前基因工程药物还只处在发展的早期，真正被正式批准投放市场的基因工程肽或蛋白类治疗药物现在还不多，不过正在开发的基因工程治疗药物有几百种，并且在迅速增加，呈现出基因工程药物的巨大潜力。基因工程药物主要包括激素、细胞因子、疫苗、抗体和寡核苷酸药物等。

(1) 基因工程激素类药物

激素是一类由生物体内分泌腺或特异性细胞产生的微量有机化合物，通过体液或细胞外液运送到特殊部位，能引起特殊的生理反应。激素在天然产生过程中的表达量是极低的，通过传统的提取方法难以获得。基因工程则可克隆其基因，使之表达获得大量产物。根据激素的化学性质可将其分为多肽蛋白类激素、甾醇类激素和脂肪酸类激素。基因工程激素类药物主要指通过基因工程方法合成蛋白质多肽类激素。目前被批准上市的有胰岛素（insulin）、人生长激素（human growth hormone）、人促卵泡激素（human follical stimulating）等。

(2) 基因工程细胞因子类药物

细胞因子是由免疫细胞和其他细胞分泌的调节生物体生理功能、参与细胞的增殖、分化和凋亡的小分子多肽类物质，称为肽类细胞因子。同激素一样，它们在生物体内只是被微量表达。表15-4是一些已上市和正在研制的基因工程细胞因子类药物。

表15-4 基因工程细胞因子类药物

名　称	作　用
各种干扰素（interferon，IFN）	抗病毒、抗肿瘤、免疫调节
各种细胞介素（interleukin，IL）	免疫调节、促进造血
各种集落刺激因子（colony stimulating factor，CSF）	刺激造血
红细胞生成素（erythropopoetin，EPO）	促进红细胞生成，治疗贫血
肿瘤坏死因子（tumor necrosis factor，TNF）	杀伤肿瘤细胞、免疫调节、参与炎症和全身性反应
表皮生长因子（epidermal growth factor，EGF）	促进细胞分裂、创伤愈合、胃肠道溃疡防治
神经生长因子（nerve growth factor，NGF）	促进神经纤维再生
骨形态形成蛋白（bone morphogenetic protein，BMP）	骨缺损修复、促进骨折愈合
组织纤溶酶激活剂（tissue-type plasminogen activator，t-PA）	溶解血栓、治疗血栓疾病
血凝因子Ⅷ、Ⅸ	治疗血友病
生长激素（growth hormone，GH）	治疗侏儒症
胰岛素（insulin）	治疗糖尿病
超氧化物歧化酶（superoxide dismutase，SOD）	清除自由基、抗组织损伤、抗衰老

(3) 基因工程抗体

抗体作为疾病预防、诊断和治疗的制剂已有上百年的发展历史。早期制备抗体的方法是将某种天然抗原经各种途径免疫动物，在血清中获得抗体，这种抗体是多种单克隆抗体的混合物即多克隆抗体。多克隆抗体的不均一性限制了对抗体结构和功能的进一步研究和应用。1975 年 Kohler 和 Milstein 首次用 B 淋巴细胞杂交瘤技术制备出均一性的单克隆抗体。杂交瘤单克隆抗体又称细胞工程抗体。利用这种技术制备的单克隆抗体在疾病诊断、治疗和科学研究中得到广泛的应用。但是，这种单克隆抗体多是由鼠 B 细胞与鼠骨髓瘤细胞经细胞融合形成的杂交瘤细胞分泌的，具有鼠源性，进入人体会引起机体的排异反应；完整抗体分子的分子质量较大，在体内穿透血管的能力较差；生产成本太高，不适合大规模工业化生产。在 20 世纪 80 年代初，抗体基因结构和功能的研究成果与重组 DNA 技术相结合，产生了基因工程抗体技术。基因工程抗体即将抗体的基因按不同需要进行加工、改造和重新装配，然后导入适当的受体细胞中进行表达的抗体分子。与单克隆抗体相比，基因工程抗体具有如下优点：①通过基因工程技术的改造，可以降低甚至消除人体对抗体的排斥反应；②基因工程抗体的分子质量较小，可以部分降低抗体的鼠源性，更有利于穿透血管壁，进入病灶的核心部位；③根据治疗的需要，制备新型抗体；④可以采用原核细胞、真核细胞和植物等多种表达方式大量表达抗体分子，大大降低生产成本。自从 1984 年第一个基因工程抗体人-鼠嵌合抗体诞生以来，新型基因工程抗体不断出现，如人源化抗体、单价小分子抗体、多价小分子抗体、某些特殊类型抗体（双特异抗体、抗原化抗体、细胞内抗体、催化抗体、免疫脂质体）及抗体融合蛋白（免疫毒素、免疫粘连素）等。抗体分子是生物学和医学领域用途最为广泛的蛋白分子。肿瘤特异性抗原或肿瘤相关性抗原、抗体独特型决定簇、细胞因子及其受体、激素及一些癌基因产物等均可以作为靶分子，目前正在进行开发和已经投入市场的抗体性药物有着广泛的用途，如器官移植排斥反应的逆转、肿瘤免疫诊断和显像、肿瘤导向治疗以及哮喘、牛皮癣、类风湿性关节炎、红斑狼疮等多种自身免疫性疾病的治疗。

(4) 基因工程疫苗

直接利用微生物制备疫苗来预防和治疗疾病已有 100 多年的历史，并且取得了巨大的成就。经典疫苗是指灭活的整个病原体或者通过不断培养获得的减毒活力病原体，它们存在着一些如含有有害成分、毒力逆转等有可能给人体带来副作用的问题。利用重组 DNA 技术发展的新型基因工程疫苗则可以将这种副作用降至最低。通过抗原基因或者含抗原表达片段的克隆、表达和纯化，人们可以制备纯化组分疫苗和亚单位疫苗等新型疫苗，这些疫苗去除了病原微生物的有害成分，安全性得到极大的提高。目前已经使用和正在研究的基因工程疫苗有多种，在细菌中包括痢疾菌苗、霍乱菌苗、结核菌苗、产毒性大肠杆菌苗等、在病毒中

包括乙肝疫苗、丙肝疫苗、艾滋病疫苗、甲型肝炎疫苗、巨细胞病毒疫苗、流行性出血热疫苗、轮状病毒疫苗等。除了上述基于蛋白质的基因工程疫苗外，自20世纪90年代发现DNA可以直接引起机体免疫应答以来，核酸疫苗成为人们极为关注的新型疫苗。核酸疫苗包括DNA疫苗和RNA疫苗，与以前的疫苗相比它具有许多优点，同时在目前也有很多有待解决的问题，其中包括其免疫机理和安全性问题，它的应用还有赖于进一步的深入研究。

3. 基因治疗

基因治疗（gene therapy）是指将外源正常基因导入靶细胞，以纠正或补偿因基因缺陷和异常引起的疾病，以达到治疗目的。也就是将外源基因通过重组DNA技术以及基因转移技术插入病人的适当的受体细胞中，使外源基因制造的产物能治疗某种疾病。基因治疗主要以两种策略达到治疗目的。其一是正常基因来纠正突变基因，也就是在原位修复缺陷基因的直接疗法，此乃理想的基因治疗策略，由于多种困难，目前尚未实现；其二是用正常基因替代致病基因的间接疗法，此法较前者难度小，也是目前众多主张采用的策略之一，并已付诸临床实践，如复合免疫缺陷综合征的基因治疗、黑色素瘤的基因治疗等。从广义说，基因治疗还可包括从DNA、RNA水平采取的治疗某些疾病的措施和新技术，例如反义技术（antisense technology）和RNA干扰（RNAi）技术。预计基因治疗在遗传病和肿瘤治疗方面将有广阔的应用前景。

二、重组DNA技术在农业上的应用

重组DNA技术应用于农业生产是农业生物技术领域中研究和开发最为活跃的部分。应用重组DNA技术将目的基因导入动、植物体内，对家畜、家禽和农作物等进行品种改良已取得了可喜的进展，尤其在农作物转基因育种方面所取得的成就更为引人注目。自从1983年首次获得转基因植物——转基因烟草以来，基因工程在农业方面的研究与应用进入了蓬勃发展阶段。目前，植物基因工程已经在多方面有了深入的发展，包括抗虫、抗病、抗除草剂、抗逆、品质改良等。此外，在植物发育调控、成熟期的调控方面也有了长足发展。基因工程在农业生产上的应用深刻影响到农业的生产方式和效益，在解决人类所面临的环境恶化、资源匮乏等问题方面初步显示出巨大的作用。由于转基因农作物品种推广和应用的时间还不长，对其可能带来的负面作用特别是对人体健康和生态环境的影响进行正确评价还需要时间。

1. 在转基因植物中的应用

(1) 利用基因工程培育抗除草剂作物

化学除草剂在现代农业中起着十分重要的作用。20 世纪 80 年代中期开始除草剂抗菌素基因的转移研究与抗性品种的开发。目前世界上采用的除草剂主要分为两大类，一类主要是通过破坏氨基酸合成途径来杀死杂草。另一类是通过破坏植物光合作用中电子传递链的蛋白来杀死杂草。现已获得的抗除草剂转基因作物有大豆、棉花、玉米、水稻、甜菜等 20 多种，抗除草剂的转基因作物占总的转基因作物的 70％以上。

(2) 利用基因工程技术改良植物品质

转基因植物育种的一个主要目标是提高农作物产量和品质。如更长久地保鲜水果和蔬菜、培育出富含有益健康的脂肪的油料作物、增强营养价值（如维他命、氨基酸、核酸）的食品、生产抗癌蛋白质含量高的大豆及其他应用。在目前的转基因作物中，涉及品质改良的占转基因植物总数的 1/5 左右，包括番茄、大豆、小麦、水稻、玉米、土豆等一些主要的粮食和蔬菜品种在内，有 30 多种转基因植物已经准许商业性种植。

(3) 利用基因工程技术培育抗虫作物

全世界粮食产量因虫害所造成的损失占 14％左右，以往长期使用化学农药防止病虫害，已经产生了严重的环境污染、生态破坏和抗药性等问题，因而抗虫基因工程在国内外受到高度重视，是植物基因工程研究和应用的热点。目前，获得了一大批转基因抗虫作物品种或品系。抗虫基因来源于细菌、植物和动物，从细菌中分离出来的抗虫基因，主要是苏云金杆菌（Bt）毒蛋白基因；从植物组织中分离出的抗虫基因，主要为蛋白酶抑制剂基因、淀粉酶抑制剂基因、外源凝集素基因等；从动物体内分离的主要是毒素基因，如蝎毒素基因、蜘蛛毒素基因等。目前应用最广的抗虫基因是苏云金杆菌的结晶蛋白的基因（Bt 毒蛋白基因)。杀虫蛋白对许多重要的农作物害虫、包括鳞翅目、鞘翅目、膜翅目、双翅目等都具有特异性的毒杀作用，而对人、畜、哺乳动物和天敌无害，因而被广泛应用于各种作物的转基因育种。目前，抗虫棉已经在国内外推广种植，但是仍然不能排除 8～10 年后出现抗性问题，以及新虫害、新病害的出现。

(4) 利用基因工程技术培育抗病作物

对植物造成重要危害的病原生物主要有真菌、细菌、病毒和线虫等，目前对前 3 种病原具有一定抗性作用的转基因作物都已培育出来，达十余种。抗病基因工程中，目前主要采用病原体的外壳蛋白基因导入植物的方法，使植物具有抗病性。例如，将 TMV、CMV、PVY、PVX、SMV 等多种病毒的外壳蛋白基因分别转入烟草、番茄、马铃薯、大豆、苜蓿等多种作物中，在不同程度减轻了病症，推迟了发病时间。

(5) 利用基因工程技术培育抗逆性强的作物

影响植物生存的主要逆境因子是干旱、高温、低温、盐胁等，抗逆基因工程主要包括抗旱、抗寒、抗热和抗盐等方面的研究。这方面工作目前尚处于起步阶段，但已取得了初步的进展。在抗寒抗冻转基因方面主要采取冷诱导基因转移、抗冻蛋白基因转移或是改善植物代谢途径中的关键酶等方法；在抗旱、抗盐碱转基因方面主要是将甘露醇、果聚糖、海藻糖等细胞渗透保护物质生物合成的关键酶基因导入作物中。

(6) 利用转基因植物生产稀有蛋白等产品

植物生物反应器将是未来基因工程发展的一个重要领域。利用植物生产口服疫苗、工业用酶、脂肪酸、药物等已逐步成为人们关注的热点和工作重心。此外，通过基因工程的方法，可以用植物生产用来制造生物塑料的底物多羟基丁酸，从而最终避免目前所谓的"白色污染"问题。植物生物反应器的研究发展使农业这一概念的外延大大拓宽，突破了传统农业范畴，延伸到工业和医药领域。

此外，重组 DNA 技术在转基因植物方面还包括：利用基因工程生产具有重要经济价值的植物次生代谢产物（生产药物、化妆品、食品添加剂等）；利用重组的植物病毒（如 MTV）感染植物，以产生大量有医用价值的蛋白；固氮基因工程；利用基因工程调控植物激素和生长发育等。

2. 重组 DNA 技术在转基因动物方面的应用

动物转基因技术是 20 世纪 80 年代初发展起来的一项新的生物技术。转基因动物研究主要表现在利用基因工程改良动物、生产药用蛋白质、动物抗病育种、建立诊断和治疗人类疾病的动物模型、生产用于人体器官移植的动物器官等方面，并已取得了初步的成就。

(1) 提高动物产品的产量和质量

1980 年 Cordon 等人用显微注射法育成带有人胸腺激酶基因片段的转基因小鼠后，1982 年 Pacmiter 等将大鼠的生长激素基因导入小鼠受精卵雄性原核中育成转基因"超级鼠"，该鼠比普通小鼠生长速度快 2～4 倍，体型大 1 倍，这一科研结果大大鼓励了该领域的科研人员。迄今为止，转基因兔、转基因羊、转基因猪、转基因牛、转基因鸡和转基因鱼等相继问世。目前，转基因动物方面的研究主要集中在：①利用外源基因提高动物生产性能，主要指的是导入各种生长激素基因以获得增重的动物。②抗冻蛋白基因、抗病蛋白基因的转移。③通过转移和表达适当的基因，培育出有突出经济性状的动物改良品种或动物新品种。

(2) 动物反应器生产有重要价值的蛋白质

用转基因动物作为生物反应器（bioreactor）开发生产新一代生物工程药物正在成为全球范围内生物高技术产业的技术制高点。其原理是将编码活性蛋白的基因导入动物的受精卵或早期胚胎内以制备转基因动物，并使外源基因在动物体内

（乳汁、血液等）高效表达，然后再分离提取目的基因编码产物，亦称为“分子农场”。对于利用转基因动物（家畜）生产转基因药物而言，最理想的表达场所是乳腺。因为乳腺是一个外分泌器官，乳汁不进入体内循环，不会影响转基因动物本身的生理代谢反应。从转基因动物乳汁中获取的基因产物不但产量高、易提纯，而且表达的蛋白经过转基因动物生物反应器表达的产物经过了充分的修饰和加工，具有稳定的生物活性，同时具有成本低、周期短、易规模化生产等优点。至今，各国科学家已制造出乳腺分泌各种医用蛋白质的转基因牛、山羊、绵羊、猪和家兔，乳腺分泌的医用蛋白质达到可商业开发水平的转基因动物已达 10 种，如 α_1-抗胰蛋白酶、人血清白蛋白、人凝血因子 IV、人凝血因子 VIII、组织性纤溶酶原激活剂（tPA）等。

（3）转基因动物生产人体器官，用于异种移植

人类同种异体器官移植近几十年来已取得了令人满意的效果，但其供体来源严重不足。而且随着人非正常死亡的逐渐减少及人寿命的延长，人供体器官将更加贫乏，异种器官移植将是解决此问题的主要途径。猪的器官大小，解剖生理特点与人类相似，组织相容性抗原 SLA 与人 HLA 具有较高的同源性，研究者普遍认为猪是人类器官移植的最理想的供体。异种移植最大的障碍仍然是移植后发生的免疫排斥反应。通过免疫排斥相关基因转基因猪的建立，对猪的器官进行遗传改造，是降低其免疫排斥反应的重要途径。现已培育出带人体基因的转基因猪，其器官移植于人体能够有效降低动物器官的排斥，这种转基因猪的成功可望解决供移植的人体器官不足的问题。

（4）克隆技术培育动物

动物克隆技术早在 20 世纪 70 年代就为科学家所关注，已有了重要进展。1997 年，英国克隆羊“多利”的产生，引起全球极大关注。“多利”是第一例经体细胞核移植出生的动物，是人类在这一领域研究中的重大突破，意味着可以快速地生产出大量的克隆动物。从本质上讲，动物克隆技术只是亚细胞水平的操作，属于细胞工程，但我们可以预期在不远的将来重组 DNA 技术将会很快应用于动物克隆，因为几乎不存在技术上的障碍。从科学技术和经济效益上分析，利用克隆技术可以加快优良家畜品种的繁育，同时对于濒危动物和家畜遗传多样性的保存具有重大意义。

三、基因工程在环境保护中的应用

微生物对环境的净化已被用于污水处理和环境净化。通过 DNA 重组可以提高某些微生物体内特异酶的活性，从而为构建能降解水中、土壤中特殊的污染化合物的微生物提供新的途径。到目前为止，已经分离到的降解性质粒有 25 个之多。将这些质粒有目的地转入到以降解有害有毒有机污染物为目的的微生物

中——即基因工程菌的构建，能大大提高降解污染有机物的效率，还有利于扩大可降解污染物的种类，有利于环保工业的发展。美国获得的速效消除水面油污的“超级菌”是美国第一个基因工程改造微生物的专利，它是将分别能降解已烷、辛烷和癸烷；降解二甲苯和甲苯；降解萘和分解樟脑的4种假单孢菌的不同质粒构建成一个质粒转入细胞内，获得了具有特殊功能的“超级菌”，其代谢碳氢化合物的活性极强，几小时便可降解掉天然菌种需1年以上才能降解的烃类物质。

第十六章　现代分子生物学的重要研究方法和技术

自20世纪50年代初沃森（J. D. Watson）和克里克（F. Crick）提出关于DNA双螺旋模型学说以及70年代初期DNA限制性内切核酸酶的发现和一整套DNA体外重组技术的建立，分子生物学技术以空前的速度蓬勃发展，建立起了一套完整的技术体系，并在生命科学的各个研究领域得到了广泛的应用，成为不可取代的研究技术手段。同时，分子生物学技术已对农、林、牧、副、渔，食品，制药，化工，石油工业以及环境监测和保护等国民经济诸多领域产生深远的影响，为人类社会带来难以估量的社会效益和经济效益。

第一节　定 量 PCR

聚合酶链式反应（polymerase chain reaction，PCR）自诞生之日起就决定了它不仅是一种高敏感、高特异的检测核酸分子的定性方法，而且也是一个能对核酸分子进行精确定量的有利工具。随着分子生物学技术研究的不断进展，定量PCR技术取得了突飞猛进的发展，不仅建立了一系列的方法，而且也诞生了许多与这些方法相匹配的新型热循环仪和实验材料。实时定量（quantitative real-time PCR，QRTPCR）技术便是一种具有革命性意义的定量PCR技术。QRTPCR是1996年由美国Applied Biosystems公司推出，由于该技术不仅实现了PCR从定性到定量的飞跃，而且与常规PCR相比，它具有特异性更强、有效解决PCR污染问题、自动化程度高等特点，目前已得到了广泛的应用。所谓QRTPCR技术是指在PCR反应体系中加入荧光基团，利用荧光信号积累实时监测整个PCR进程，最后通过标准曲线对未知模板进行定量分析的方法，同时也可以用于RT-PCR测定mRNA的量来确定基因转录水平的变化。

一、常用的实时定量PCR方法

1. TaqMan探针技术

定量PCR技术自产生以来，不断发展完善，到目前为止已经非常成熟了。标记方法由最初单一的染料法，发展到了特异性更高的探针法，如TaqMan、Molecular Beacon、Fret等不同方法的标记探针，其中由于TaqMan探针的广泛使用，QRTPCR的方法也称为TaqMan法。TaqMan探针法是高度特异的定量PCR技术，其核心是利用*Taq*酶的5′→3′外切核酸酶活性，切断探针，产生荧光信

号。由于探针与模板是特异性结合，所以荧光信号的强弱就代表了模板的数量。在 TaqMan 探针法的定量 PCR 反应体系中，包括一对 PCR 引物和一条探针。探针只与模板特异性地结合，其结合位点在两条引物之间。探针的 5′端标记有报告基团（reporter，R），如 FAM、VIC 等，3′端标记有荧光淬灭基团（quencher，Q），如 TAMRA 等。当探针完整的时候，报告基团所发射的荧光能量被淬灭基团吸收，仪器检测不到信号。随着 PCR 的进行，Taq 酶在链延伸过程中遇到与模板结合的探针，其 5′→3′外切核酸酶活性就会将探针切断，报告基团远离淬灭基团，其能量不能被吸收，即产生荧光信号（图 16-1）。所以，每经过一个 PCR 循环，荧光信号也和目的片段一样，有一个同步指数增长的过程。信号的强度就反映出模板 DNA 的拷贝数。

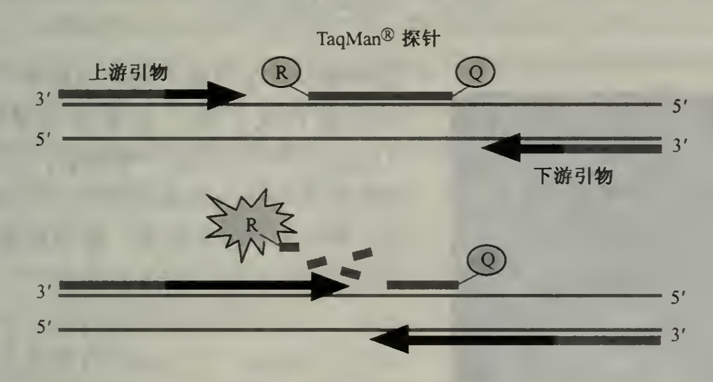


图 16-1　TaqMan 探针技术原理示意图

TaqMan 探针根据其 3′端标记的荧光淬灭基团的不同分为两种：普通的 TaqMan 探针和 TaqMan MGB 探针。MGB 探针的淬灭基团采用非荧光淬灭基团（non-fluorescent quencher），本身不产生荧光，可以大大降低本底信号的强度。同时探针上还连接有 MGB（minor groove binder）修饰基团（图 16-2），可以将探针的 T_m 值提高 10℃左右。因此为了获得同样的 T_m 值，MGB 探针可以比普通 TaqMan 探针设计得更短，既降低了合成成本，也使得探针设计的成功率大为提高。因为在模板的 DNA 碱基组成不理想的情况下，短的探针比长的更容易设计。实验证明，TaqMan MGB 探针对于富含 A/T 的模板可以区分得更为理想。

2．SYBR Green I 荧光染料技术

SYBR Green I 是一种只与 DNA 双链结合的荧光染料（图 16-3）。当它与 DNA 双链结合时，发出荧光；从 DNA 双链上释放出来时，荧光信号急剧减弱。因此，在一个体系内，其信号强度代表了双链 DNA 分子的数量。SYBR Green I 荧光染料法定量 PCR 的基本过程是：①开始反应，当 SYBR Green I 染料与 DNA 双链结合时发出荧光。②DNA 变性时，SYBR Green I 染料释放出来，荧光

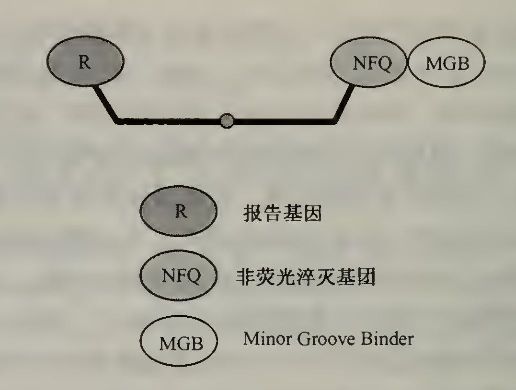


图 16-2　TaqMan MGB 探针

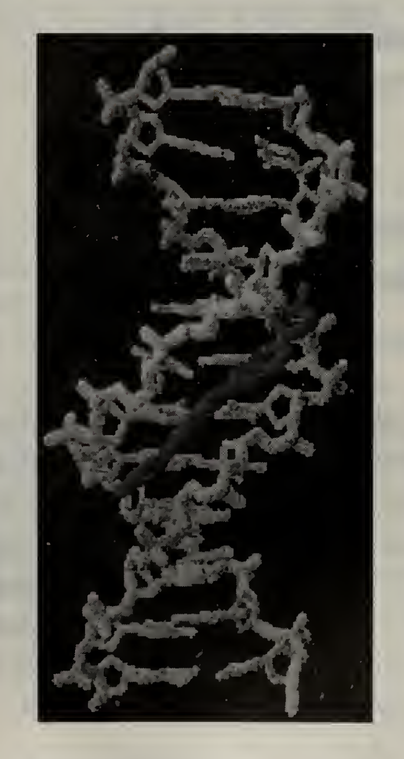


图 16-3　SYBR Green I 荧光染料与 DNA 双链的结合

急剧减少。③在聚合延伸过程中，引物退火并形成 PCR 产物。④聚合完成后，SYBR Green I 染料与双链产物结合，定量 PCR 系统检测到荧光的净增量加大。SYBR Green I 是一种 DNA 结合染料，能非特异地掺入到双链 DNA 中去。在游离状态下，它不发出荧光，但一旦结合到双链 DNA 中以后，便可以发出荧光。它的最大优点就是可以与任意引物、模板相配合，用于任意反应体系中，从经济角度考虑，它也比其他的探针的价格要便宜得多，但由于它能与所有的双链 DNA 结合，所以一旦反应体系中出现非特异扩增，那它就会影响到定量结果的可靠性与重复性。

3. 实践中的问题

QRTPCR 已被广泛运用于分子生物学的各个领域，就目前的应用情况来看，虽然取得了很好的效果，但其在方法学上的选择、敏感性问题、重复性问题等都一直是争论不休的。

(1) 方法的选择

研究者在实验中往往想要得到目的基因的绝对定量，因为绝对定量对目的基

因的表达差异有直接的反映。绝对定量最为简单的方法就是标准曲线法，用一系列已知浓度的标准品制作标准曲线，在同等条件下目的基因测得的荧光信号量同标准曲线进行比较，从而得到目的基因的量。该标准品可以是纯化的质粒dsDNA，体外转录的RNA，或者是体外合成的ssDNA。目的基因与标准品在不同的反应管内同时进行扩增，使用的荧光材料可以是杂交探针、水解探针或是SYBR Green I。有研究表明，这一方法的动力学范围可以达到9个数量级，远比终点法的要高。这一方法虽然简便，但是为了得到可靠的实验结果，有两个先决条件必须得到满足，一是所选用的标准品必须与目的基因的扩增效率一致，要达到这个要求，需要标准品同目的基因的同源性尽可能的高，不仅长度相似，而且其碱基组成也尽量相同；二是要尽可能地优化实验条件和模板的准备过程。就标准品而言，实践中有外参照和内参照之分，作为外参照的标准品与目的基因不在同一反应体系中进行扩增，所以它无法检测也不可能补偿可能出现在目的基因反应体系中的影响因素，如PCR抑制因素。事实上，在绝大多数的实验中，要想使标准品和目的基因的扩增效率绝对完全一致是不可能的。为了弥补单独使用一个标准品作为外参照的不足，研究者便引入一个内参照同目的基因在同一反应体系中同时进行扩增，以反映体系中可能出现的PCR影响因素，此即所谓的竞争性PCR。竞争性PCR所使用的内参照具有与目的基因序列相同的引物结合位点，但不具有相同的探针结合位点。

(2) 重复性问题

重复性是判断实验结果优劣的重要指标，其主要判断指标为标准差（SD）和变异系数（CV）。对于特定的分析系统，重复性的高低决定了PCR反应能检测出样本中目的基因的初始浓度变化的最小值。由于PCR反应本身具有不精确扩增的特性，它所造成的结果精确性要远小于经典的临床化学检验和免疫分析。对于绝大多数实验，一般都可以计算出所设计体系能检测出样本初始浓度的最小值，根据多次实验可以求得该最小值的变异情况和它的可信区间范围，变异越小或是可信区间范围越小，说明这一反应体系的精确度越高。在QRTPCR过程中，从样本的准备到扩增，再到定量的进行，实验中的每个方面都与实验的重复性息息相关，除了加样的准确性和实验所选用的仪器固定参数的限制之外，影响结果重复性较为关键的因素还包括：①PCR反应扩增的效率，如果在反应体系中扩增效率不一致，就会使目的基因在单位时间内的产量发生差异，从而影响到结果的稳定，要解决这个问题，必须尽量优化实验条件，使反应体系达到最佳扩增效率；②目的基因的初始浓度，初始拷贝数越低，结果的重复性越差，为了保证获得精确的结果，应使用初始浓度具有较高数量级的样本，如果待测样本中目的基因的量处于反应体系的检出限附近，那么最好是使用多管重复以保证结果的可靠性；③标准曲线的影响，对于必须进行绝对定量的研究，标准曲线是必不可少的，虽然标准品和样本之间的差异始终存在，但是制作一个好的标准曲线对定量

结果至关重要，在制作标准曲线时，应至少选择5个稀释度的标准品，涵盖待测样本中目的基因量可能出现的全部浓度范围，理想的标准品应与样本具有高同源性，最好是选择纯化的质粒DNA或是体外合成、转录的RNA（用于RT-PCR），在制备过程中，应使用规范的步骤或是根据所购买的试剂盒提供的标准操作手册进行。

（3）灵敏度问题

QRTPCR由于使用了荧光物质作为定量的工具，使得其灵敏性又大大地提高了，有人报道QRTPCR的灵敏性要高出传统的终点定量PCR大约250倍。在理论上，只要设计的引物对目的基因有足够的特异性，PCR应可检测出只含一个拷贝目的基因的样本，也的确有人报道在特定的反应条件下，QRTPCR可以检测出单细胞水平的基因组DNA，但是在大多数情况下，对基因组DNA而言，它的最低检出限一般在pg到fg级，对于病毒、质粒而言，其检出限一般在10^2拷贝数以上。影响QRTPCR灵敏性的因素众多，除了对一般PCR反应均存在的影响因素如反应体系、*Taq*酶的活性之外，QRTPCR还有其特殊的影响因素。

4．影响QRTPCR的因素

（1）反应体系中形成的引物二聚体的影响

引物二聚体是非特异性退火和延伸的产物，它的形成不仅会影响到扩增的效率，而且由于SYBR Green I可以与所有的双链DNA结合，所以会在反应体系中出现特异性产物与引物二聚体竞争SYBR Green I的现象，从而降低了实时PCR的敏感性。

（2）循环数

一般的QRTPCR反应只须25～30个循环便可获得满意的结果，但是对于那些极微量的待测样本而言，适当增加循环数可以提高反应的检出限，有文献报道当循环数从25增加到34个循环时，QRTPCR的最低检出限可提高到10^2个模板分子。但是并非循环数增加得越多，其敏感性就会越高。

（3）Mg^{2+}的浓度

Mg^{2+}的浓度将影响到QRTPCR的敏感性，推荐使用的浓度为3mmol/L。Mg^{2+}浓度对敏感性的影响主要存在于两方面，首先，Mg^{2+}是影响Taq酶活性的关键因素，如果Mg^{2+}的浓度无法达到使Taq酶发挥最佳活性，无疑将会影响到实时定量的敏感性；其次，Mg^{2+}的浓度过高，会增加引物二聚体的形成，从而导致敏感性降低。所以在反应中选择合适的Mg^{2+}浓度条件，是相当重要的。

5．C_t值的含义

在QRTPCR技术中，有一个很重要的概念——C_t值。*C*代表cycle，t代表threshold，C_t值的含义是：每个反应管内的荧光信号达到设定的域值时所经历

的循环数。荧光域值（threshold）的设定可以通过 PCR 反应前 15 个循环的荧光信号作为荧光本底信号，荧光域值的缺省设置是 3～15 个循环的荧光信号的标准偏差的 10 倍。

研究表明，每个模板的 C_t 值与该模板的起始拷贝数的对数存在线性关系，起始拷贝数越多，C_t 值越小。利用已知的起始拷贝数的标准品可做出标准曲线，因此，只要获得未知样品的 C_t 值，就可以通过标准曲线计算出该样品的起始拷贝数。

6. 怎样确定 C_t 值?

实验操作中，C_t 值定义为在基线上方产生可检测到的统计学上显著的荧光发射时所对应的 PCR 循环次数（图 16-4)。“基线上方”也就是阈值高度的量化定义，是基线范围内荧光信号强度标准偏差的 10 倍。阈值所在的横线与 PCR 扩增曲线的交点所指的 PCR 循环次数就是 C_t 值。基线范围的定义是从第 3 个循环起到 C_t 值前 3 个循环止，其终点要根据每次实验的具体数据调整，一般取第 3 到第 15 个循环之间。早于 3 个循环时，荧光信号很弱，扣除背景后的校正信号往往波动比较大，不是真正的基线高度；而在 C_t 值前 3 个循环之内，大多数情况下荧光信号已经开始增强，超过了基线高度，都不宜当做基线来处理。

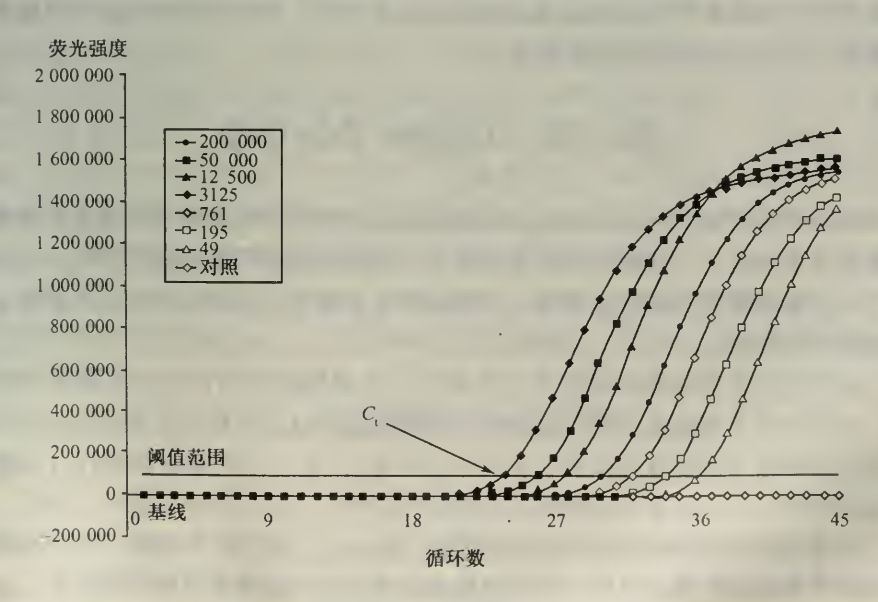


图 16-4　C_t 值和阈值

7. 为什么终点定量不准确?

我们都知道理论上 PCR 是一个指数增长的过程，但是实际的 PCR 扩增曲线

并不是标准的指数曲线，而是S形曲线。这是因为随着PCR循环的增多，扩增规模迅速增大，*Taq* 酶、dNTP、引物，甚至DNA模板等各种PCR要素逐渐不敷需求，PCR的效率越来越低，产物增长的速度就逐渐减缓。当所有的 *Taq* 酶都被饱和以后，PCR就进入了平台期。由于各种环境因素的复杂相互作用，不同的PCR反应体系进入平台期的时机和平台期的高低都有很大变化，难以精确控制。所以，即使是重复实验，各种条件基本一致，最后得到的DNA拷贝数也是完全不一样的，波动很大。

8. 标准曲线、重复实验和阴性、阳性对照

定量实验，误差是不可避免的。设立重复实验，对数据进行统计处理，可以将误差降低到最小。所以定量实验的每个样本至少要重复3次以上，严格的定量更应当重复6～8次，以满足小样本统计的要求。

如果做绝对定量，则标准曲线需要在5个点以上。标准曲线使用的标准品是浓度已知的DNA样本，可以自己制备，也可以购买商品化的试剂盒。其PCR反应条件应当与未知样本的一致，以便在同一反应板上同时定量。

阴性对照中不加模板DNA，而以水或缓冲液代替，用于检验是否存在PCR污染。阳性对照则用于检验PCR试剂和实验操作上可能出现的问题。如果实验中设立了阳性内对照（internal positive control，IPC）IPC，则IPC既可以用来校正数据，也可以起到阳性对照的作用。

第二节　核酸分子的杂交

核酸分子杂交（moleucular hybridization）技术是分子生物学领域中最常用的基本技术方法之一。其基本原理是把具有一定同源性的两条核酸单链在一定的条件下（适宜的温度及离子强度等）可按碱基互补原则退火形成双链。此杂交过程是高度特异性的。

杂交的双方是待测的核酸序列及探针。待测核酸序列可以是克隆的基因片段，也可以是未克隆化的基因组DNA和细胞总DNA。将核酸从细胞中分离纯化后可以在体外与探针直接杂交（膜上印迹杂交），也可以在细胞内进行（细胞原位杂交）。

用于检测的已知的核酸片段称为探针（probe）。为了便于检测，探针必须用一定的手段加以标记，以利于随后的检测。常用的标记物是放射性同位素。近年来发展了一些非放射性标记物，如生物素、荧光素、地高辛等，检测这些标记物的方法是极其灵敏的。

一、核酸分子杂交的原理

DNA 分子杂交实质上是双链 DNA 的变性和具有同源性序列的两条单链的复性过程。

(一) 变　　性

加热、有机溶剂及低盐浓度等因素都可以导致 DNA 二级结构破坏，双螺旋解旋、两条链完全解离，但是未破坏其一级结构，此过程称为 DNA 的变性。DNA 的变性伴随着 DNA 一系列理化性质的变化，如黏度降低、旋光性的改变及 260nm 紫外吸收值增加等。

(二) 复　　性

变性的两条互补单链，在适当的条件下重新缔合形成双链的过程称为复性（renaturation）或退火（annealing）。热变性的 DNA 溶液在低于 T_m 值 25℃ 的温度下维持相当长的一段时间，则可使之恢复到天然的双链结构状态。复性的速度受到诸多因素的影响。① DNA 浓度：DNA 浓度直接影响到 DNA 单链间碰撞的几率，DNA 浓度越大，复性速度越快。② DNA 的分子质量：大分子质量的 DNA 扩散速度较慢，也难于形成正确配对，因此复性速度较慢。③ 温度：温度过高，有利于 DNA 变性而不利于复性；而温度过低，则少数碱基配对形成的局部双链不易解离，难以继续寻找正确的配对，适宜的复性温度是较 T_m 值低 25℃。④ 离子强度：离子强度过低不利于复性，常采用 0.15～1.0mol/L 盐溶液进行复性研究。⑤ DNA 分子的复杂性：DNA 总量一定时，基因组越复杂，其中特定序列的拷贝数就越少，互补序列的浓度就越低，因而复性反应速度越慢。

二、固相支持物的选择

固相支持物的种类很多。一种良好的固相支持物应具备以下几个特性：①具有较强的结合核酸分子的能力，一般要求每平方厘米结合核酸分子的量不低于 10μg，最好能达到数十微克；②与核酸分子结合后，应不影响其与探针分子的杂交反应；③与核酸分子的结合稳定牢固，能经受杂交、洗膜等操作过程而不至于脱落或脱落很少；④非特异性吸附少，在洗膜条件下能将非特异性吸附在其表面的探针分子洗脱掉；⑤具有良好的机械性能，如柔软性好、韧性强等，以便于操作。

（一）硝酸纤维素滤膜

硝酸纤维素滤膜（nitrocellulose filter membrane）具有较强的吸附单链 DNA 和 RNA 的能力，特别是在高盐浓度下，其结合能力可达 80～100$\mu g/cm^2$。吸附的单链 DNA 或 RNA 经真空烘烤后，依靠疏水性相互作用而结合在硝酸纤维素膜上。另外，硝酸纤维素膜还具有杂交信号本底较低的优点。虽然硝酸纤维素膜是应用的最为广泛的一种固相支持物，但它存在一些难以克服的缺点。首先由于硝酸纤维素膜是依靠疏水作用结合 DNA，这种结合并不十分牢固，随着杂交及洗膜的进行。DNA 会慢慢脱离硝酸纤维素膜，特别是在高温情况下，从而使杂交效率下降。另外，硝酸纤维素膜质地较脆，操作不方便，须十分小心。所以现在更多的人倾向于使用尼龙膜。

（二）尼　龙　膜

尼龙膜（nylon membrane）是目前较理想的一种核酸固相支持物。尼龙膜结合单链及双链 DNA 或 RNA 的能力较硝酸纤维素膜更强，可达 350～500$\mu g/cm^2$。而且经烘烤或紫外线照射后，核酸分子可以牢固地结合在尼龙膜上，特别是经短波紫外线照射后，结合更加牢固。另外，由于韧性较强，操作起来很方便；对于小分子的核酸片段亦有较强的结合能力，甚至短至 10bp 的核酸片段它也能结合；同时尼龙膜还可以重复使用。其缺点就是杂交信号本底较高，但是可以通过加大预杂交液中的特异性封闭试剂的方法加以克服。

三、印迹技术

印迹技术是指将待测核酸分子结合到一定的固相支持物上的方法，这些结合在固相支持物上的核酸分子即可与存在于液相中的探针分子进行杂交。选择良好的固相支持物与有效的转移方法是此项技术成功与否的两个关键因素。

印迹的方法很多，根据核酸种类的不同，可以分为 Southern 印迹法和 Northern 印迹法。

（一）Southern 印迹

Southern 印迹是指将电泳分离的 DNA 片段转移到一定的固相支持物上的过程。DNA 分子被限制酶酶切，经琼脂糖凝胶电泳将所得 DNA 片段按分子质量大小分离，然后将含 DNA 片段的琼脂糖凝胶变性，并将其中的单链 DNA 片段转移到硝酸纤维素膜或其他的固相支持物上，而各 DNA 片段的相对位置保持不

变。这种滤膜可用于下一步的杂交反应。利用 Southern 印迹可进行克隆基因的酶切图谱分析、基因组中某一基因的定性及定量分析、基因突变分析及限制性片段长度多态性分析等。

（二）Northern 印迹

Northern 印迹是指将 RNA 变性及电泳分离后，将其转移到固相支持物上的过程，从而用于杂交反应以鉴定其中特定 mRNA 分子的大小与含量。Northern 印迹的基本原理与 Southern 印迹相同，但 RNA 变性方法与 DNA 不同，不能用碱变性，因为碱会导致 RNA 的水解，需用甲醛变性。与 DNA 相比，RNA 操作具有更高的难度。这主要是因为 Rnase 极其活跃，而且十分稳定，在 100℃ 煮沸 10min 活力没有明显丧失，因此必须严格注意到 RNase 对 RNA 的降解作用。

（三）斑点及狭缝印迹

将 RNA 或 DNA 变性后直接点样于固相支持物上，用于基因组中特定基因及其表达的定性及定量分析研究，称为斑点或狭缝印迹（dot and slot blot），也是实验室常用的技术之一。与 Southern 和 Northern 印迹法相比，其优点是简单、迅速；可在同一张膜上同时进行多个样品检测；对于核酸粗提样品的检测效果也较好。其缺点是不能鉴定所测基因的分子质量，而且特异性不高，有一定比例的假阳性。

第三节　基因沉默与基因剔除

研究基因沉默可帮助人们进一步揭示生物体基因遗传表达调控的本质，在基因工程中克服基因沉默现象，从而使外源基因能更好的按照人们的需要进行有效表达；利用基因沉默在基因治疗中有效抑制有害基因的表达，达到治疗疾病的目的，所以研究基因沉默具有极其重要的理论和实践意义。

一、基因沉默

基因沉默（gene silencing）是指生物体中特定基因由于种种原因不能表达，是研究基因功能的重要手段。基因沉默现象是导致基因不能正常表达的重要因素之一，其作用机制主要有三种：位置效应（position effect）、转录水平的基因沉默（transcriptional gene silencing，TGS）和转录后水平的基因沉默（post-transcriptional gene silencing，PTGS）。根据目前的知识，重复序列是基因沉默的普遍

诱因，甲基化是基因沉默的直接原因。

位置效应是指基因在基因组中的位置对其表达的影响。外源基因进入细胞核后首先整合到染色质上，其整合位点与表达有密切的关系。如果整合到甲基化程度低、转录活性高的常染色质上其表达受两侧 DNA 序列的影响。植物基因组常是由具有相似 GC 含量的 DNA 片段相互嵌合在一起的，外源基因的插入打乱了它们的正常的组合。生物体可以通过外源基因与其两侧序列 GC 含量的差别来识别外源基因，然后激活甲基化酶，使外源序列甲基化而降低其转录活性。

转录水平上的基因沉默是 DNA 水平上调控转录的结果，主要是由启动子甲基化或导入基因异染色质化所造成的。二者都和转基因重复序列有密切关系。自身甲基化外源基因如果以多拷贝的形式整合到同一位点上，形成首尾相连的正向重复或头对头、尾对尾的反向重复，则不能表达。而且拷贝数越多，基因沉默现象越严重。

转录后基因沉默是 RNA 水平基因调控的结果，比转录水平的基因沉默更普遍。PTGS 表现为一种序列特异性的 RNA 降解过程，主要作用于同源性较高的转录产物，包括具有同源性的内源基因的转录产物。它广泛存在于各种生物中，直到最近人们才发现在植物中被称为共抑制（cosuppression）、真菌中被称为抑制（quelling）、动物中称为 RNA 干扰（RNA interference，RNAi）的现象可能具有很大的相似性，都与 PTGS 现象有关。现在普遍认为 PTGS 并非一种偶然现象，而是生物在长期进化过程中所形成的对病毒、转座因子和其他可转移核酸的防御系统，因为这些外源核酸的引入很可能使宿主细胞内的平衡机制遭到致命破坏，而转基因沉默，无非是宿主将外源基因视为对自己有害的序列而将其抑制。细胞对外源核酸的这种抑制作用具有序列特异性的特点，所以一旦细胞内转录后沉默机制被启动，细胞对转入核酸的同源序列就具有了“免疫”能力。目前对 PTGS 机制的了解仍不是十分清楚。这是因为生物在长期的进化过程中防御外来基因侵入的机制本身具有多样性，所以很难用一种理论或学说来解释基因沉默现象（图 16-5）。

RNAi 现象是广泛存在于生物中的一种普遍现象。研究显示，生物体内有一套 RNA 监视系统，可以由多种异常 RNA 来激发。如果外来核酸是 DNA（包括转基因、重组基因、DNA 病毒、扩增子等），靶标 RNA 需要在细胞核中完成转录后运转到细胞质中，而侵入细胞质的病毒 RNA 可以直接提供靶标 RNA。各种不同的靶标 RNA（包括与外源基因同源的内源基因和外来 DNA 产生的 RNA 以及病毒的 RNA）由寄主的 RdRP（依赖于 RNA 的 RNA 聚合酶）或病毒自身的 RdRP 通过多种不同的途径把靶标 RNA 转变成双链 RNA，从而通过 RNAi 引发的 PTGS。PTGS 被引发后就不再需要 RdRP。关于双链 RNA 介导的 RNAi 特异性靶标 RNA 的降解，有研究认为生物体内存在着一种复合酶：RNAi 核酸酶，该酶具有双链 RNA 结合、RNase 和 RNA 解旋酶三个活性区。首先双链 RNA 结

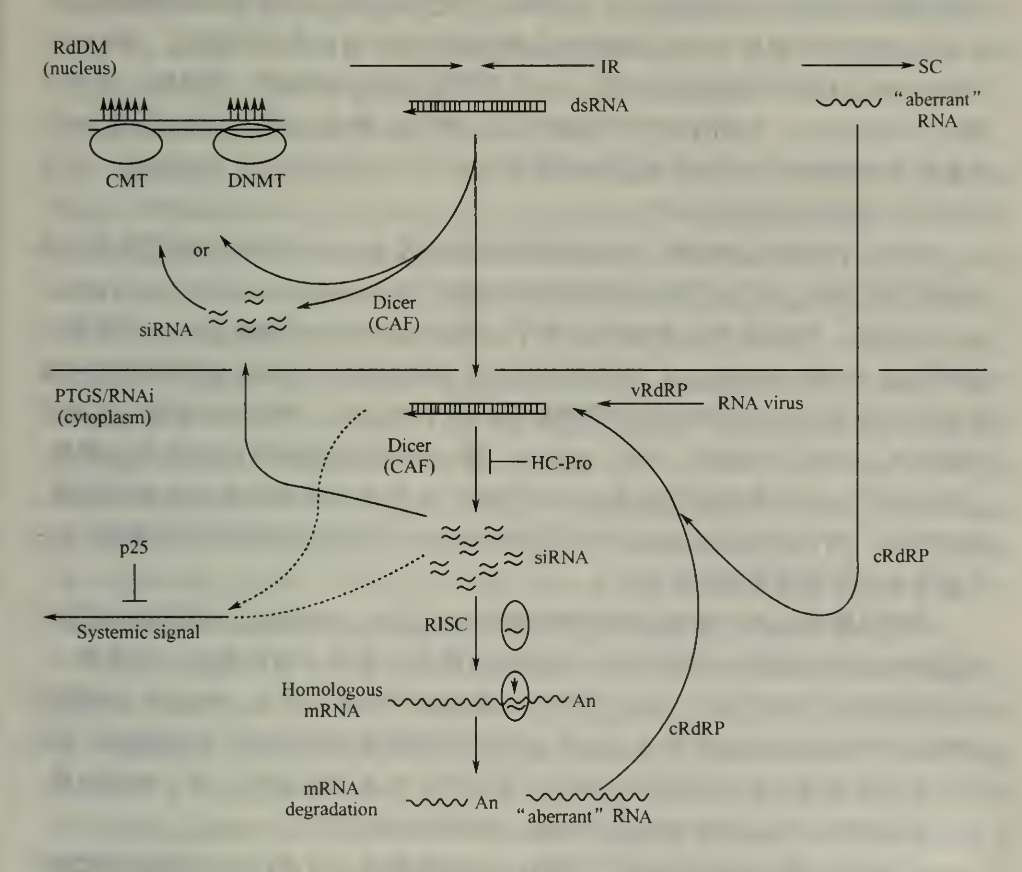


图 16-5 基因沉默原理示意图

Dicer：一种能促进 RNA 成熟和降解的酶。

siRNA：short interfering RNA，小干扰 RNA。

stRNA：small temporal RNA，小瞬时 RNA。

Genome：基因组。

Ago：一种蛋白质的名称。

RISC：RNA-induced silencing complex，RNA 沉默诱导复合体。

RdRP：RNA-dependent RNA polymerase，依赖 RNA 的 RNA 聚合酶。

adapter：接头分子。

dsRNA：双链 RNA。

inverted repeat transcripts：反转序列转录。

non-coding RNA genes：非编码 RNA 基因。

transgene arrays：转基因阵列。

transposon：转座子。

translational arrest：翻译停滞。

degrade：降解。

genomic modification：基因组修饰。

Transgene：转基因。

合到该酶的双链 RNA 结合区并引导该酶识别靶标 RNA，接着该酶的解旋酶区完成 ATP 依赖性的靶标 RNA 与该酶结合的双链 RNA 的正义链的换位，RNase 在靶标 RNA 结合位点附近完成切割，从而使靶标 RNA 能被进一步降解，产生大量的小片段 RNA，包括序列特异性的约 25ntRNA。载有序列特异性的双链 RNA 的游离复合酶再去识别并降解其他的靶标 RNA，产生更多的约 25ntRNA，从而使 PTGS 具有持久性和系统性。

转录水平上的基因沉默、转录后基因沉默以及 RNAi 引起的基因沉默都与基因的同源性有关，称为同源依赖性的基因沉默（homology-dependent gene silencing，HDGS）。TGS 是发生在核内的事件，而 PTGS 发生在细胞质中。两者都与超甲基化（hypermethylated）或形成的 RNA 双链有关，TGS 现象中甲基化主要发生在启动子区域，基因的转录受到抑制；而 PTGS 中，甲基化主要发生在基因的编码区，基因能够转录，产生 mRNA，但 mRNA 在细胞质中被特异地降解。二者在时间上并非简单的前后关系，在空间上也不因为核膜的存在而相互隔离。研究发现，PTGS 可能以 RNA 的方式影响 TGS，而且越来越多的研究表明 TGS 可能与 PTGS 有着相似的机制。

基因沉默现象具有重要的理论和实践意义。首先，在理论上如果能阐明基因沉默的机理对于提高转化效率和外源基因的表达活性具有指导意义；在实践上，由于植物基因工程的兴起，使人们的许多愿望有可能成为现实，但是由于转基因的转化效率较低，稳定性不高、表达水平难以预料这些缺点限制了植物基因工程的发展，转基因沉默的研究的实践意义首先在于提高表达活性，减少基因沉默。另外基因沉默还可以用于研究基因功能、基因的表达调控。

克服基因沉默已经成为基因工程的一个重要课题，目前针对上述基因沉默的机制，初步提出了以下一些对策：①由于重复或同源序列是基因沉默的普遍诱因，所以在构建表达载体时，应尽量使得所设计的序列与内源序列的同源性较低，以减少或避免配对。另外，选用外源基因插入基因组中拷贝数低的，最好是单拷贝的转基因植株，亦可减少重复序列的存在。②甲基化是基因沉默的直接原因，转基因甲基化的程度与基因沉默的程度成正相关，人们正在试图在载体上加上有去甲基化功能的序列以防止甲基化。③实验表明转基因的侧翼核基质结合序列（matrix attachment region，MAR）会在一定程度上避免位置效应，提高基因的表达。MAR 具有限定 DNA 环的大小，使之成为相对独立的结构功能单位的作用。

二、基因剔除

基因剔除技术是在基因定位整合技术以及小鼠胚胎干细胞体外培养的基础上建立的一种新兴的生物技术。该技术使转基因技术更前进了一步，为人们在动物

体内分析某一特定的基因提供了有力的手段，这一最新技术最先用于动物模型的复制上。1988 年人类培育了第 1 例基因剔除动物模型，揭开了基因剔除动物模型建立的序幕。进入 20 世纪 90 年代，在生命科学研究的各个领域如免疫学、人类疾病模型复制和基因功能鉴定等方面都有着广泛的应用。

基因剔除（gene knock-out）就是用基因转移方法将一段 DNA 转移至受体细胞内，转移基因可以取代受体细胞染色体基因组上一段长度相同的 DNA 片段，取代方式有随机组合和同源重组两种，同源重组即用转移的一段 DNA 序列取代基因组上相应的一段序列。基因剔除可以用突变的基因剔除相应的正常基因，也可以用正常基因剔除相应的突变基因。目前主要采用突变基因（mutated gene）敲除正常基因以产生定位突变（target mutation）。首先选定需要突变的基因的全部或部分 DNA 序列，通过插入、修饰、删除或置换等手段使其突变，克隆到载体中建成靶载体，将靶载体导入小鼠 ES 细胞（胚胎干细胞）中，使之与细胞染色体内同源的靶序列进行重组，达到定位突变细胞内该基因的目的。然后在细胞水平和分子水平筛选富集发生突变的细胞，并注射到小鼠囊胚腔内，将这些囊胚导入假孕母鼠子宫中。所产生的子代雄性嵌合鼠与正常雌鼠交配可获得携带该突变基因的纯合鼠。最后对纯合鼠的表型、基因型及基因表达产物进行分析。

采用同源重组定位突变的基因剔除系统为探讨基因的生理功能和表达调控机制提供了有效的方法。这种方法的最大优点是它能对所有已经克隆的基因进行突变，且能对同一基因进行多种形式的突变。目前采用的基因剔除是使该基因功能全部丧失。今后的研究可使这一作用进一步扩大与完善：①可敲除基因的一小段序列观察其功能细节，也可以缺失大片段研究成簇基因的功能与调控机制；②从多基因调控机制入手，使得突变体额外获得某一基因的功能，部分或完全丧失某一基因功能；③除突变结构基因外，也可敲除调控元件以及编码反式调节因子的基因。深入研究上述不同类型的突变体，将逐步揭示多种基因的功能与表达调控机制。

在分析基因剔除实验结果时，需注意一个普遍存在的现象。当某一基因被敲除后，该基因在正常状态下特异表达的组织中并非都出现纯合突变鼠的表型。这一现象的可能原因是：①常规形态检查法难以测出组织细胞的微细变化；②该基因产物对其所表达的组织并非绝对必要；③可能存在其他的基因的功能补偿了突变基因的缺陷。另外，对分化与发育起重要作用的同源异形基因，成簇排列且在进化上高度保守，成簇特点在基因表达调控中显然有重要的意义。但其中单个基因的表达系统似乎也有相应的功能，与上述推断相矛盾。其可能的原因是：①由于进化上高度保守，转基因动物中已有相应的遗传背景；②同源异型基因的典型瞬间表达不易在基因剔除系统中反映出来。因此不管是实验设计还是结果分析，均应考虑系统的整体性与系统诸要素之间的相关性。对于一个小系统的反常现象，不可以简单地看做新发现的证据。

第四节　基因表达差异分析

生物的表型主要是由其内在的基因表达所决定的。高等生物以前估计一般具有 1×10^5 个基因，不同细胞间基因表达的差异决定了生命活动的多样性，如发育与分化、内环境稳定、细胞周期调节等。正常代谢或各种病理改变，不管是单基因突变或是受多基因体系调控，本质上都是基因表达的改变所致。因此，分析正常细胞与突变细胞之间基因表达的差异，可以为生物生命的活动过程提供重要的信息。传统上研究基因表达的方法如 Northern 印迹或斑点杂交，既费时又费力。随着分子生物学及其技术的飞速发展，出现了大量的新方法、新技术研究基因表达的差异。按其技术特点可分为三类：①以杂交为基础的技术，包括 Northern 印迹、DNA 微阵列（基因芯片）；②以 PCR 为基础的技术，如差异显示、代表性差异分析；③以测序为基础的技术，如表达序列标签（EST）、基因表达系列分析（SAGE）等。下面对其中有代表性的技术作一介绍。

一、传统研究基因表达差异的方法

Northern blot 即用同位素或其他非放射性的物质标记的 DNA 或 RNA 探针与吸印在滤膜上的被分析样品总 RNA 杂交，根据杂交信号判定已知基因在被分析样品中表达与否或表达量的技术。近 30 年过去了，该技术在分子生物学领域仍有极为广泛的应用，任何基因的表达都要用它来证实。另外一项比较基因表达差异的技术是核酸酶保护分析（RNase-protection assay）。用标记好的被检测基因与样品总 RNA 杂交，由于单链特异的核酸酶不能降解 mRNA/DNA 杂交双链，用核酸酶降解后电泳，就能检测到被保护的也是被检测基因对应的 mRNA 的存在。这些技术的一个共同点就是只能对已知的基因进行鉴别分析。

削减 cDNA 文库构建技术（subtractive cDNA library）使基因表达差异分析的比较有了质的飞跃。用过量参比材料 cDNA 与被分析材料总 mRNA 杂交，分离不能形成杂交双链的 mRNA 并用于构建 cDNA 文库，就能分析出那些在被分析材料中特异表达的全部基因。该技术最重要的改进是差异表达基因的选择扩增。把被分析材料（tester）的双链 cDNA 用限制性内切核酸酶消化，连上有生物素末端标记的接头序列，用过量的参比样品（drive 或 control）cDNA 与之杂交，用单链 DNA 酶降解不能复性的接头序列，在生物素帮助下分离被分析样品特异 cDNA 片段的杂交双链，再经 PCR 扩增，就能获得大量的被分析材料中特异表达基因的片段。这是因为只有被分析样品特异表达的基因才能形成两端都有接头序列的双链 DNA 而使接头不被降解，从而能被接头序列引物有效扩增。

二、差异展示

差异展示（differential display）是一种方便、快捷、能一次比较多个样品基因表达差异的技术。由于灵敏度极高，发育时期的微小不同步、微量组织细胞不一致等都能造成大量的带纹差异，从而使真正的差异表达基因的鉴别非常困难。同时，污染的基因组 DNA 也能因随机引物的使用在 PCR 扩增过程中产生带纹，造成严重的干扰，因此样品必须用 DNA 酶处理，以获得高纯度的 RNA。

该技术的基本流程是：在反转录酶作用下，用三条与 mRNA 3′端配对的锚定引物合成 cDNA，然后用锚定引物和随机引物的组合进行 PCR 扩增，在序列分析胶上带纹的有无和强弱就是不同材料中基因表达差异的体现。把有差异的 cDNA 带洗脱、再扩增、克隆和测序，就能分离出大量差异表达基因的不完整序列并可以作为进一步分离基因的探针。在鉴别和分离发育生化相关基因和疾病形成相关基因等方面，差异展示获得了极成功的应用。该技术具有能同时对多个样品进行相互比较且需要的 RNA 起始量非常微小等优点。其缺点就是假阳性较多，但是，尽管如此，差异展示技术仍然是比较基因表达差异的最有效方法。

三、代表差异分析
（representational difference analysis，RDA）

RDA 最初是为分离基因组特有的 DNA 序列而设计的，后来被用于基因表达差异分析。它结合了 PCR 技术和削减杂交技术的优点，能极有效地分离在一个 cDNA 群体中存在而在另一个群体中不存在的基因序列，但对那些在表达量上有差异的基因的鉴别和分离却较困难。一般期望被分析样品含有特异表达的基因，但经常做交互分析，即先确定一个样品为被分析样品，然后交换，把参比样品作为被分析样品再做一次分析，把分析样品和参比样品的 mRNA 群体反转录成 cDNA，再用限制性内切核酸酶消化，并在 cDNA 限制性内切核酸酶消化片段两端连上接头序列，经 PCR 扩增，就能获得足够量 cDNA 群体的限制性片段，把接头序列从 cDNA 片段的两端用酶切掉，再给被分析样品加上新接头序列，对照的切割接头后则不加新接头，这样就只有被分析样品的 cDNA 片段两端有外加的接头序列。用 1∶100 比例混合的被分析样品和参比样品 cDNA 片段作模板进行 PCR 扩增，来源于对照的 DNA 片段由于没有新接头，不能和引物退火因而不能被扩增，而测试的和对照都有片段，过量的参比样品 cDNA 片段在复性时很快与被分析样品的相同序列杂交而阻碍扩增，只有被分析样品中特有片段获得了有效地扩增。RDA 的最大优点是能富集被分析样品特异表达的 cDNA 片段。为了达到更好的富集效果，通常要进行第二轮甚至第三轮的杂交 PCR 操作，这样大大

降低了电泳分析的假阳性条带。

四、基因表达系列分析 (serial analysis of gene expression, SAGE)

SAGE 是通过比较序列信息来比较基因表达差异的技术。它能用少数几个测序反应获得不同的组织或细胞基因表达差异的全部信息。SAGE 包括每个表达基因独有的 10 到 14 个碱基标签序列的获得，能用于克隆和测序的基因标签序列的组装和标签种类以及频率的统计三个基本步骤。根据标签的种类和频率就能推论各标签对应基因的表达情况。

基因表达系列分析是一个以测序为基础的基因组工具，具有发现新基因以及定量基因表达分析的功能。SAGE 技术，是 1995 年由约翰·霍普金斯大学的 Velculeseu 等提出的，此方法用来鉴定哪些基因被表达以及量化表达水平。SAGE 方法的两个基本原理是：①需要一个短的寡核苷酸标记物（SAGE 标记物，10bp），它必须包含充分的信息能特异性地代表一个转录本；②以一个串联体的形式将 SAGE 标记物随机地串联起来，使得测序经济高效。具体解释如下：

最初估计人类基因组中大约有 100 000 个不同的基因，一个 10bp 大小的寡核苷酸序列有 1 000 000 多种不同的组合方式。如果在每个转录本中特定的区域能获得一个 10bp 的标记物，那么此标记物足以特异性的代表这一转录本。第二条原则是对短的核苷酸序列标签随机线形串联组成的模板（concatemer，串联体）进行测序，这些短核酸标记物来源于不同的转录本。这个串联体的测序结果将提供多种基因的信息。

SAGE 作为一种强有力的全新的方法，能够对基因表达的整体模式做出综合而又全面的评价。其操作过程简要介绍如下：首先，用多聚 T 柱子分离 poly (A)$^+$RNA，用生物素标记的引物 T18，以多聚（A)$^+$RNA 为模板合成被生物素标记的 cDNA，再被一种“锚定酶”切割，被生物素标记的 cDNA 片段与包被有链亲和素的珠状物结合。含有标记酶识别位点的两种寡核苷酸（linker）分别连接到这种结合性 cDNA 片段的 3′终端，这样形成一个寡核苷酸 + 3′cDNA 标记物复合体。标记酶使寡核苷酸 + 3′cDNA 标记物复合体从珠状物上消化下来，形成寡核苷酸 + 标记物复合体，复合体的 3′末端被纯化。带有两种不同寡核苷酸（linker）的短标记物可互相连接形成 102bp 大小的双标记物。将包含有两个标记物（SAGE 双标记物）和两个 linker 结合的双标记物进行 PCR 扩增。Linker 结合的 SAGE 双标记物被标记酶（II 型限制性内切核酸酶）从识别下游约 20bp 处切开产生大约 26～28 个碱基对的寡核苷酸聚合物（SAGE 双标签），经纯化、连接形成串联体，克隆到 M13 载体中，扩增 DNA 并测序。利用 SAGE 软件对测序结果进行分析能够将标记物分类和量化，这些标记物反应了一个已知细胞系或

组织中转录本的特性和数量。

五、生物芯片技术

生物芯片（biochip）技术是近年来发展起来的新型实用技术，已成为高效、大规模获取相关生物信息的重要手段。所谓生物芯片就是缩小了的生物化学分析器，通过芯片上微加工获得的微米结构与生物化学处理结合，将成千上万个与生命相关的信息集成在一块厘米见方的氧化硅玻璃或塑料等材质制成的芯片上，实现对化合物、蛋白质、核酸、细胞或其他生物组分准确、快速、大信息量的筛选或检测。基因芯片，又称 DNA 微阵列（DNA microarray），是一种最重要的生物芯片，广泛用于基因表达差异分析、新基因发现、基因突变及多态性分析、基因组文库作图、疾病诊断、药物筛选、基因测序等。

基因芯片是利用原位合成法或将已经合成好了的一系列寡核苷酸以预先设定的排列方式固定在固相支持介质表面（硅片、玻片、尼龙膜等），形成高密度的寡核苷酸的阵列，以用于杂交分析。其中，寡核苷酸的长度与链内互补序列及 T_m 值等因素的限制，一般在 25bp 以内，针对不同的用途，长度不一。目前有关基因芯片的来源和介绍都很多，但大多数为公司化的产品，尚无明确统一的标准来限定基因芯片的质量。

基因芯片技术的突出特点在于高度并行性、多样性、微型化和自动化。高度并行性不仅可大大提高实验的进程，并且有利于基因芯片技术所展示图谱的快速对照和阅读。多样性是指在单个芯片中可以进行样品的多方面分析，从而大大提高了分析的精确性，避免了因不同实验条件产生的误差。微型化是当前芯片制造中普遍的趋势，其好处是可以减少试剂用量和减少反应液体积，从而提高样品浓度和反应速度。高度自动化则可以降低制备芯片的成本和保证芯片的制造质量不易波动。

所有的基因芯片技术都包含五个基本要点：生物学问题的提出和 DNA 方阵的构建、样品的制备、杂交反应、杂交图谱的检测、数据的处理及建模。

基因芯片的制备方法大致可以分为点样法和在片合成法。在基因芯片的使用过程中常常需要同时对大量不同的基因片段进行扩增和标记。因此，要求对 PCR 引物的选择、扩增条件以及 PCR 产物的标记进行摸索。基因芯片杂交结果的检测有多种方法，目前主要为荧光标记方法。根据获得的荧光图谱，进行数据的分析，建立相应的数据库。

将基因芯片技术应用于基因表达差异的研究的做法是：把不同条件下某生物体内转录出来的所有 mRNA 经标记成为探针后，再与代表它所有基因而制成的寡核苷酸方阵杂交，通过分析杂交位点及其信号强弱就可得出不同情况下每个基因是否表达及表达多少。

蛋白质芯片又称蛋白质微阵列（protein microarray），其制作原理类似于基因芯片。不同的是，蛋白质芯片所用的样品是提纯了的蛋白质、多肽或从 cDNA 表达文库中提取的蛋白产物。它是继基因芯片之后，作为基因芯片功能的补充发展起来的。与基因芯片的概念类似，它是在一个基因芯片大小的载体上，点布高密度的不同种类的蛋白质，然后再用标记了荧光染料或其他的标记物标记的已知抗体或配体一起同芯片上蛋白质竞争结合，在扫描仪上读出荧光强弱，计算机分析计算出待测样品结果。在此基础上可发展到对各种蛋白质、抗体以及配体的检测，弥补了基因芯片检测的不足。

第五节　DNA 启动子的活性研究

在 DNA 分子上，调控蛋白结合以及转录起始复合物装配地方的一段 DNA，一般位于转录起始点上游，称为启动子。要分析启动子的活性，就必须将假定的启动子序列插入到报告基因的上游，为了选择要插入报告载体的合适 DNA 片段，必须首先确定翻译起始密码子和转录起始位点的位置。哺乳动物基因的转录起始位点通常位于翻译起始密码子上游 50～200bp 处。但是，通过序列同源性鉴定到了可能的 TATA 框和其他可能的调控元件，就简单地推断启动子在 ATG 密码子的上游是不正确的。可以在目的基因旁侧序列存在的条件下检测报道活性。但是在未得到转录起始位点图谱的情况下，仅利用灵敏的报道分析检验时，就难以区分检测到的活性是由真正的启动子驱动，还是由能驱动低水平转录并存在于整个基因组中的隐蔽调控元件驱动产生。因为启动子区总是位于转录起始位点周围，如果转录起始位点已经明确，结果的可靠性就会大大增加。

目前，研究启动子活性的方法比较多，下面主要介绍采用氯霉素乙酰基转移酶试验（CAT 实验）。

一、CAT 实验原理

自从 Gorman 等 1982 年首次介绍以氯霉素乙酰基转移酶（chloramphenicol acetyltransferase，CAT）作为报告分子以来，*CAT* 基因就被广泛用作许多哺乳动物细胞标准培养系的报告基因。CAT 满足报告分子所需的许多标准。

CAT 由原核基因编码，可以产生对抗生素氯霉素的抗性。CAT 可以把酰基基团从酰基辅酶 A（acetyl-CoA）加到氯霉素上，从而使抗生素结合原核核糖体并抑制其功能的活性丧失。因为哺乳动物细胞中没有 *CAT* 基因，而且 CAT 酶活性分析具有高灵敏度和可重复性，所以原核 *CAT* 基因成为广泛应用于真核启动子分析的报告基因。实验中通常把一段真核 DNA 放在 *CAT* 基因编码区的 5′上游，得到的质粒再去转染哺乳动物细胞。转染的基因表达一段时间后再检测

CAT 活性。如果实验设置正确，CAT 活性值与转染质粒的转录水平之间会有线形关系。对假定的启动子 DNA 片段作突变，转染后测定突变体合成 CAT 的活性，有可能以此来确定 DNA 中的调控元件。尽管 CAT 实验分析具有可定量的特性，但要记住本方法并不直接测量转录情况，某些参数如 CAT 蛋白在转染细胞中的半衰期会影响结果。还有一个潜在的问题，CAT 活性可能由起始于质粒中异常位置的转录物产生，这种异常转录可能使对所测试启动子活性的估计有很大程度的夸大。有些情况下，可以将转录终止信号放在感兴趣的启动子的上游，可以消除假转录造成的背景。如果用虫荧光素酶作报告基因，检测比 CAT 方便，但稳定性和重复性不如 CAT。

二、CAT 实验方法

1. 转染细胞沉淀的制备

1）将经过转染并在组织培养皿中生长的单层细胞轻轻地吸掉培养液。单层细胞用不含钙盐和镁盐的 PBS 洗 3 次；

2）把培养皿以一定角度放置 2～3min，使最后残留的 PBS 流到一侧。吸掉 PBS 残液。每个培养皿加 1ml PBS，用橡胶棒把细胞刮倒离心管中。冰浴，直到所有的培养皿处理完毕；

3）室温下以最大速度离心 10s 收集细胞。用 1ml 冰冷的 PBS 轻轻地重悬细胞沉淀，然后再次离心收集细胞。去掉细胞沉淀与管壁中的 PBS 残液。细胞沉淀可以储存在 −20℃ 以便将来分析用。

2. 制备细胞抽提物

裂解细胞是用反复冻融法或用含有去污剂的缓冲液处理细胞。后一种方法快速简便，而且适用于在 96 孔板上测量 CAT。

3. 用薄层层析法测定 CAT 活性

1）取 50μl 细胞抽提物在 65℃ 孵育 10min 使内源性的去乙酰基酶失活。如果这时抽提物混浊不透明，在 4℃ 以最大速度离心 2min 以去除微粒；

2）每个待测样品用 80μl CAT 反应混合液混合，37℃ 孵育。孵育时间的长短与细胞抽提物中 CAT 的浓度有关，而 CAT 的浓度又与所研究的启动子强度和细胞类型有关。在大多数情况下，孵育 30min 到 2h 就足够了；

3）每个样品加 1ml 乙基乙酸，振荡 3 次，每次 10s，彻底使溶液混匀。混合物在室温下以最大速度离心 5min；

4）用移液管把 900μl 有机相转移到新管中，操作时应小心，避免水相和中

间相混入。把含有下层相的管子作为放射性废物丢弃；

5）把管子放入旋转蒸发器，真空下使乙基乙酸挥发约 1h；

6）把 10～15μl 已溶解的反应产物加到硅胶薄层层析板的起始处。板上的起点用软石墨铅笔标记。每次加样 5μl，用吹风机把样品吹干；

7）准备盛有 200ml 薄层层析溶剂的薄层层析槽。把薄层层析板放入槽中，关上容器，然后使溶剂前沿移动到板上端大约 75%的地方；

8）把薄层层析板从槽中取出来，在室温下干燥。用放射性墨水在薄层层析板上做点标记以便于对齐板与胶片的位置，然后用板对 X 射线胶片曝光，或者把板放在磷屏分析的盒子内，盒子在室温下放置适当的时间；

9）冲洗 X 射线胶片并与薄层层析板校对位置。此外，还可以用层析图对磷屏分析设备的图像板曝光或把薄层层析板拿去扫描；

10）为了对 CAT 活性定量，需要从薄层层析板中把放射性斑点切割下来，然后用液闪计数器测定放射性量。取另一份细胞抽提物（来自于步骤中的③），用 Bradford 分析法等快速比色法确定抽提物中的蛋白浓度。测定蛋白浓度前，把 Triton X-100 的浓度稀释为≤0.1%，以免发生交叉反应。CAT 活性表示为每单位时间每毫克细胞抽提物蛋白中产生乙酰基化产物的摩尔数。

第六节　DNA 与蛋白质相互作用研究方法

一、凝胶阻滞分析——Gel-shift 实验

1. Gel-shift 实验原理

当 DNA 片段与蛋白质结合在一起形成 DNA-蛋白质复合物后，DNA 片段的分子质量及所带的电荷发生改变，因而在聚丙烯酰胺凝胶体系中的电泳迁移率也发生变化，由此得名为电泳迁移率改变分析法，由于在一般情况下 DNA-蛋白质复合物的迁移率会变慢，所以一般称之为凝胶阻滞分析法（gel retardation assay），或称为 Gel-shift 实验。此方法具有高度灵敏性，可以检测出 10^{-15}mol 的 DNA 结合蛋白。采用该方法可以鉴定目的基因的调控序列中是否存在特定的 DNA 结合蛋白的结合位点，也可以用来检测特定的细胞核蛋白是否存在某种基因的 DNA 结合蛋白，因此常用于基因转录调控因子的初步筛选，是研究序列特异性 DNA 结合蛋白的常用方法。

2. 操作步骤

1）凝胶制备：

40%丙烯酰胺溶液	5ml
10%过硫酸铵	0.5ml

10×TAE　　　　　　　　　　　5ml

混匀后，加入 TEMED10μl，混匀，灌入电泳板中。

2）预电泳：将凝胶在 1×TAE 中于 4℃条件下预电泳 30～60min，电泳条件为 11 V/cm，在电泳过程中，用蠕动泵循环上、下槽的缓冲液。

3）结合反应：取细胞核提取液 1～4μl 加入到 eppendorf 管中；加结合缓冲液至总体积为 12μl；10mmol/L $MgCl_2$ 和 125mmol/L 亚精胺各 2mol/L；加入多聚（dI·dC）3μl，混匀，室温放置 2～3min；加入用［^{32}P］末端标记的 DNA 片段（1ng）1μl。

4）30℃保温 30min 后，点样于预电泳后的凝胶中，用 30mA 恒定电流进行电泳。

5）游离 DNA 片段至凝胶下端时，停止电泳。

6）电泳完毕，将凝胶转移到滤纸上，用凝胶干燥器真空加热干燥后，进行放射自显影。

3．注意事项

1）DNA 片段的长度在 200bp 左右为宜，DNA 片段太长，在聚丙烯酰胺凝胶电泳中不易分离，电泳迁移率的轻微改变难以被检出。

2）为检查 DNA 与核蛋白结合的特异性，需要进行竞争性抑制试验，其方法是向反应体系中加入过量的非标记的特异性 DNA 片段。如果结合是序列特异性的，则这些非标记的 DNA 片段可与标记的 DNA 片段竞争性地与核蛋白结合，其结果是造成放射自显影 X 底片上的滞后带消失。用这种方法，将多种人工合成的双股寡核苷酸或较短的 DNA 限制酶切片段分别加入反应体系中进行竞争抑制试验，可将核蛋白在 DNA 分子上的结合位点进行初步定位，因为只有相当于结合位点 DNA 序列的寡核苷酸片段才能抑制它们的结合。

3）反应体系中加入多聚（dI·dC）的目的是抑制核蛋白与 DNA 片段的非特异性结合，起封闭作用。在核蛋白提取液纯度较低的情况下，多聚（dI·dC）的用量要进行摸索，也可以用鲑精 DNA 或小牛胸腺 DNA 代替多聚（dI·dC)。

4）核蛋白的用量对实验结果具有明显影响，蛋白量过高则非特异性结合增加，过低则有可能出现假阴性结果。核蛋白的用量与其对 DNA 的亲和性、核蛋白提取液中特定 DNA 结合蛋白的丰度以及反应条件等多种因素有关。因此，在正式试验前，应对核蛋白量进行摸索，以能形成清晰的滞后带，又有适量的游离 DNA 带残存为宜。

5）亚精胺和 $MgCl_2$ 的浓度也显著影响核蛋白与 DNA 的结合，其用量也需要经过预试验摸索。

6）DNA 片段的用量一般为 1.0ng，但可视情况适当增减，以能形成清晰的滞后带而游离 DNA 带又不过浓为宜。

7) 为保持 DNA-蛋白质复合物在凝胶中的稳定性，电泳缓冲液一般采用低离子强度的 0.25×TBE 或 1×TAE。但是，低离子强度的缓冲液缓冲能力不强，长时间的电泳会导致正负极缓冲液的 pH 发生变化，因此，宜将正负极的缓冲液进行不间断的循环。

8) 为除去凝胶中的不纯物质，在标本进行电泳分离之前应在稳定电流情况下进行预电泳至电压不再上升为止。电泳一般在 4℃条件下进行为宜。

二、DNase I 足迹法

DNase I 足迹法（foot printing）是在体外鉴定 DNA 结合蛋白在 DNA 分子上的结合位点的实验方法。这种方法的基本原理是 DNA 分子上的特定序列与 DNA 结合蛋白结合后，DNA 结合蛋白可保护该部位的 DNA 序列不受 DNase I 的攻击，也即将待测双链 DNA 片段中的一条单链的一端用 ^{32}P 标记后与核蛋白进行结合，之后加入适当浓度的 DNase I 进行部分消化（平均一个 DNA 分子上产生一个切口），经变性后用测序聚丙烯酰胺凝胶电泳分离，放射自显影，即可出现以相差一个核苷酸为阶梯的 DNA 条带。因为结合有蛋白质的序列不能被 DNase I 所切断，所以不能检出 DNA 条带。因此，将与核蛋白结合和未与核蛋白结合的 DNA 片段分别用 DNase I 消化，以 Maxam 和 Gilbert 化学测序法处理的 DNA 片段作为 Marker 同时进行电泳，经比较两者的 DNA 条带，即可确定 DNA 分子上的蛋白质结合位点。

在放射自显影图谱上，DNA 梯度条带在相应于 DNA 结合蛋白的结合部位中断形成一个空白区域，好像蛋白在 DNA 上留下的足迹，因而被形象地称为足迹法。实际上，足迹是一个统计学的结果，因为 DNase I 切断所形成的每一个 DNA 片段本身并不能反映 DNA 结合蛋白的结合部位，而是反映此部位不是结合位点，只有将所有 DNA 片段综合在一起后才能反映出精细的结合位点。

三、甲基化干扰足迹法

甲基化干扰足迹法是利用硫酸二甲酯（DMS）使 DNA 分子中的 G^7 及 A^3 位发生甲基化，当蛋白结合位点上的 G 和 A 被甲基化后，便会干扰 DNA 结合蛋白与结合位点的结合反应从而形成足迹。与 DNase I 足迹法相比，甲基化干扰足迹法的优点是：①DNase I 足迹法需要较高的蛋白浓度以使所有的结合位点被饱和，而本方法不受非完全结合的影响，因为 DNA 与蛋白质结合后需经凝胶电泳迁移率改变分析，将与蛋白结合的 DNA 探针与游离探针分离；②蛋白质与 DNA 的结合是动态反应，在蛋白从 DNA 上的结合位点解离时，DNase I 可导致结合位点的消化，而本方法则不会，因为凡能结合蛋白质的 DNA 在结合位点上均不

会有甲基化修饰；③本方法的分辨率较 DNase I 足迹法高。本方法的缺点是 DMS 的甲基化修饰作用具有碱基特异性，主要修饰 G^7 位和 A^3 位，不修饰 T 和 C，因此，当结合位点的一条单链上不存在 G 和 A 时，则需要测定另一条互补单链，综合两条链的结果才能得出精确的结合位点。

第七节　蛋白质与蛋白质的相互作用

蛋白质间的相互作用存在于机体细胞的生命活动过程中，生物学的许多现象如复制、转录、翻译、剪切、分泌、细胞周期调控、信号转导和中间代谢等均受到蛋白质之间相互作用的调控。有些蛋白质由多个亚基组成，他们之间的相互作用就显得更为普遍，有些蛋白质结合紧密，而有些蛋白质只有短暂的相互作用，然而不论哪种情况，它们均控制着大量的细胞活动事件，如细胞的增殖、分化和死亡。通过蛋白质之间的相互作用，可改变细胞内蛋白质的动力学特征；也可以产生新的结合位点，改变蛋白质对底物的特异性；还可以失活其他的蛋白质，调控其他基因表达。因此，研究蛋白质与蛋白质之间的相互作用的意义重大，其检测方法的也备受重视。目前研究蛋白质与蛋白质相互作用的方法有生化方法（如蛋白质亲和层析、亲和印迹、免疫沉淀及交联）以及发展到现在的分子生物学方法（如以基因文库为基础的蛋白质探测、噬菌体展示技术以及双杂交系统）。通过这些方法的联合使用，由实验得出的蛋白质间相互作用的结论更为可靠。分子生物学方法尤其是双杂交系统的建立，使研究蛋白质间的相互作用变得更为简捷。

一、双杂交系统

双杂交系统（two-hybrid system）又称相互作用陷阱（interaction trap）是近几年来建立的研究蛋白质相互作用的新方法。其基本原理是：构建二个基因重组杂交蛋白，一个杂交蛋白为 DNA 结构域（A）与感兴趣的蛋白质（X）融合而成，表示为 A+X，另一个杂交蛋白为 DNA 转录激活结构域（B）与待筛选的蛋白（Y）融合而成，表示为 B+Y，在酵母细胞中含有两个报告基因，通常为 *Leu* 和 *lacZ* 基因，这两个基因受一段 DNA 序列调控，通常为 DNA 上游激活序列（upstream activating sequence，UAS）或 *LexA* 操纵基因，它们可被 DNA 结合结构域结合。报告基因的检测灵敏度为 *Leu* 大于 *LacZ*。将两个杂交蛋白重组体转化到酵母细胞中后，若 X 与 Y 可相互作用，则转录激活结构域通过 DNA 结合结构域与 UAS 或 *LexA* 的结合激活报告基因的转录，结果酵母菌能在缺乏 Leu 的培养基上生长，同时在含有 X-gal 的培养基上呈蓝色。反之则否。挑取阳性酵母菌，分离含有编码 B+Y 的质粒，进行序列分析，其中 Y 基因编码的 Y 蛋白

即为新发现的与 X 相互作用的蛋白质。

二、噬菌体展示技术

1. 噬菌体展示技术的基本原理

噬菌体展示技术（phage display）的基本原理是将外源基因插入噬菌体信号肽序列与主要衣壳蛋白基因Ⅷ或基因Ⅲ之间，构建成融和蛋白噬菌体库，表达的融和蛋白可以呈现在噬菌体表面，而不影响噬菌体的完整性。利用呈现的多肽与感兴趣的蛋白亲和结合的性质可以高效率的筛选所需的基因，即先将感兴趣的蛋白固相化，如结合在塑料板上，加上文库噬菌体并与感兴趣的蛋白反应。经过洗涤，无结合力的噬菌体被洗掉，有亲和力的噬菌体即被俘获，洗脱的噬菌体可以再感染大肠杆菌而繁殖，这样“吸附—洗涤—洗脱—繁殖”的富集过程称为淘筛(panning)。一轮淘筛可使噬菌体富集 10^3 倍，经过几轮淘筛就能从 $10^9 \sim 10^{10}$的噬菌体文库中筛选得到与感兴趣蛋白相互作用的蛋白。

2. 用于噬菌体展示的载体

在噬菌体展示系统中，外源 DNA 片段插入到基因 III 或基因 VIII 中位于切割部位下游的几个核苷酸处。用重组病毒 DNA 转染的 *E. coli* 合成并分泌单一成分的感染性“融合噬菌体”颗粒，其表面展示有外源 DNA 编码的氨基酸。因为在转染的细菌噬菌体颗粒表面的每一份 pIII 或 pVIII 拷贝均载有外源 DNA 编码的序列，因此它们是以紧密装配的多价方式来展示。

针对特定的目的，已经建立了特殊的噬菌体展示载体。其中最著名的是含有一个琥珀（UAG）链终止突变的噬菌体载体，这一突变位于外源性 DNA 插入片段的下游，pIII 或 pVIII 结构的上游。当重组噬菌粒用来转化 *E. coli* 的非抑制性菌株时，由外源 DNA 编码的蛋白质在琥珀密码子位置终止，并分泌进入培养基中。当采用载有免疫球蛋白基因片段的噬菌粒时，从各个抑制基因阴性的转化子的上清液中，可以筛选到能够与抗原结合的可溶性抗体片段。然而，当噬菌粒用来转化表达琥珀抑制子的细胞时，可以合成整个融和蛋白，抗体片段按正常方式展示在所分泌的噬菌体颗粒的表面。

3. 肽的噬菌体展示

将序列不同而长度相同的合成寡核苷酸插入基因 III 或基因 VIII 中，就可以生成巨大的、可以展示数千万个肽的重组噬菌体文库。因为编码了目标的基因是细菌噬菌体基因组的一部分，那么通过其蛋白质产物的亲和选择，细菌噬菌体展示就可以用来分离特定的 DNA 序列。获益于噬菌体肽展示的广泛的生物系统，

包括抗原表面定位、蛋白质-蛋白质相互作用分析以及抑制剂、激动剂和拮抗剂的分离。

4. 肽展示文库的构建

建立肽展示文库的步骤原理很简单：合成正确长度的随机寡核苷酸，采用PCR转换成双链分子，用限制酶消化，连接到已经在细菌噬菌体基因III或VIII中的适当位置构建好的相应的限制位点上。连接产物转染到*E. coli*，并收集经培养产生的噬菌体。许多噬菌体展示载体，比如像fUSE系列的载体都载有四环素抗性基因，这可以在含有抗生素的培养基上对转染细胞作为克隆选择和维持。然后，噬菌体基因组就可像质粒一样繁殖。这种选择减少了多轮感染期间对pIII功能的要求。细胞在液体培养基中生长几个小时后，可以随时收集噬菌体。

三、免疫共沉淀技术

当细胞在非变性条件下被裂解时，完整细胞内存在的许多蛋白质-蛋白质结合保持下来。这一事实可被用于检测和确定生理条件下相关的蛋白质-蛋白质间相互作用。如果蛋白质X的抗体免疫沉淀X，那么与X在体内结合的蛋白质Y可能也沉淀下来。基于与蛋白质X的生理性相互作用，蛋白质Y的免疫沉淀被归类为免疫共沉淀。这种方法最常用于测定两种目标蛋白质是否在体内结合，但也用于确定一种特定蛋白质的新的作用搭档。

两种蛋白质间生理性相互作用的最严格证实，也许是它们从细胞提取液中的免疫共沉淀。假定的相互作用可能首先用有效的高通量方法确定，如酵母双杂交筛选，然后用免疫共沉淀证实两种蛋白质是真正的生理性相互作用。通过用编码相关蛋白质的质粒瞬间转染细胞，然后表达蛋白质，常可加快以此方法对蛋白质-蛋白质相互作用的检测。然而，从质粒的过表达生成的蛋白质可能驱动那些非生理性的两种蛋白质间的相互作用。因此，证实两种蛋白质在它们的自然的非转染细胞内的相互作用常是更好的选择。

用免疫共沉淀探询与已知蛋白质相互作用的新蛋白质，是确定完整细胞内生理性相互作用的有效方法。此方法的主要缺点是费时费力，并且需要大量培养细胞，以及可能检测不到低亲和力的瞬间的蛋白质-蛋白质相互作用。另外，免疫共沉淀仅对从细胞溶出的、并存留在生理复合物中的蛋白质有效。因此，免疫共沉淀可能不适合于检测构成巨大的、不溶性的大分子结构的蛋白质-蛋白质相互作用。

主要参考文献

桂建芳. 2002. 发育生物学. 北京:科学出版社

杨俊杰. 1998. HOX基因与造血细胞发育. 国外医学生理,病理科学与临床分册,18(4):325~327

Agrawal R K et al. 1998. Visualization of elongation factor EF-G on the *E. coli* ribosome: the mechanism of translocation. Proc. Natl Acad Sci USA, 95: 6134~6138

Anderson W F et al. 1981. Structure of the Cro repressor from bacteriophage λ and its interaction with DNA. Nature, 290: 754

Angel P, Allegretto E A, Okino S T, Hattori K, Boyle W J, Hunter T, Karin M. 1988. Oncogene jun encodes a sequence-specific transactivator similar to AP-1. Nature, 332: 166~171

Ashkenazi A, Dixit V M. 1998. Death recceptors : Signeling and modulation. Science, 281:1305~1308

Ausubel F M et al. 1995. Short protocols in molecular Biology. 3rd ed. New York: John Wiley & Son, Inc

Avery O T, C M MacLeod, M McCarty. 1944. Studies on the chemical nature of the substance inducing transformation of pneumococcal types. J Exp Med, 79: 137

B Eramhill, A Kornberg. 1988. Cell, 52 : 743

Bass B L, T R Cech. 1984. Specific interaction between the self-splicing RNA of Tetrahymena and its guanosine substrate: implications for biological catalysis by RNA. Nature, 308: 820

Bell S, T A Baker. 1998. Polymerases and the replisome: machines within machines-Cell, 92: 295~305

Bellin T, Pulz M, Matussek A, Hemphen H G, Gunzer F. 2001. Clin Microb, 39:370~374

Benjamin Lewin. 1997. Genes v1. Oxford: Oxford University Press

Berchtold H, Reshetnikiva H C O A et al. 1993. Crystal structure of active elongation factor Tu reveals major domain rearrangements. Nature, 365:126~132

Berger, J S Gamblin, S Harrison, J Wang. 1996. Strucure and mechanism of DNA topoisomerase II. Nature, 379: 225~232

Bianchi M E, C M Radding. 1983. Insertions, deletions, and mismatches in heteroduplex DNA made by RecA protein. Cell, 35: 511

Bienz M. 1985. Transient and developmental activation of heat-shock genes. Trends Biochem Sci, 10: 157

Blackwood E M, J T Kadonaga. 1998. Going the distance a current view of enhancer action. Science, 281: 60~63

Burley S K, R G Roeder. 1996. Biochemistry and structural biology of transcription factor IID(TFIID). Ann Rev Biochem, 65: 769~799

C D Webb et al. 1997. Cell, 88: 667

C Vassylyev, K Morikawa. 1997. Curr Opin Struc Biol, 7 : 103

Calos M P, J H Miller. 1980. Transposable elements. Cell, 20: 579

Chu G. 1997. Double strand bread repair. J Biol Chem, 272: 24097~24100

Clapham D E, Neer E J. 1993. New roles of G Protein β γ-dimers in transmembrane signaling. Nature, 365: 403~406

Cox M M, I R Lehman. 1987. Enzymes of general recombination. Ann Rev Biochem, 56: 229~262

Cozzarelli N. 1980. DNA gyrase and supercoiling of DNA. Science, 207: 953

Cross M, Dexter T M. 1991. Growth factors in development, tranformation, and tumorigenesis. Cell, 64: 271~

Darnell J E Jr. 1997. STATs and gene regulation. Science, 277: 1630～1635

Divecha N, Irvine R F. 1995. Phospholipid signaling. Cell, 80: 269～278

Dutta A, S Bell. 1997. Initiation of DNA replication in mammalian cells. Ann Rev Cell Devel Biol, 13: 293～332

Engels W R. 1983. The P family of transposable elements in Drosophila. Ann Rev Genetics, 17: 315

Fan Z, Beresford P J et al. 2003. Tumor suppressor NM23-H1 is a granzyme A-activated DNase during CTL-mediated apoptosis, and the nucleosome assembly protein SET is its inhibitor. Cell, 112 : 659～672

Fishel R, T Wilson. 1997. MutS homologs in mammalian cells. Curr Opin Genet Devel, 7: 105～113

Frank D N, N R Pace. 1998. Ribonuclease P : unity and diversity in a tRNA processing ribozyme. Ann Rev Biochem, 67: 153～180

Frank J. 1998. How the ribosome works. Ann Scientist, 86: 428～439

Freistroffer D V et al. 1997. Release factor RF3 in *E. coli* accelerates the dissociation of release factors RF1and RF2 from the ribosome in a GTP-dependent manner. EMBO J, 16: 4126～4133

Giordano A, Romano G. 1999. Gene therapy-where are we going? Clin Ter, 150:3

Goeddel D V. 1991. Gene expression technology. USA: Academic Press Inc

Gough N. 1981. The rearrangement of immunoglobulin genes. Trends Biochem Sci, 6: 203

Hamm J, A I Lamond. 1998. Spliceosome assembly: the unwinding role of DEAE-box proteins. Curr Biol, 8: 532～534

Haseltine W A. 1983. Ultaviolet light repair and mutagenesis revisited. Cell, 33: 13

Hooykaas P J J, R A Schilperoort. 1985. The Ti plasmid of Agrobacterium tumefaciens: a natural genetic engineer. Trends Biochem Sci, 10:307

James M, Abdullah, et al. 2001. Cloning and characterization of *hepp*, a novel gene expressed preferentially in hematopoietic progenitors and mature blood cells. Blood Cells Molecules and Diseases, 27(3): 667～676

Landy A, W Ross. 1977. Viral integration and excision: structure of the lambda att site. Science, 197: 1147

Little J W, D W Mount. 1982. The SOS regulatory system of *E. coli*. Cell, 29: 11

M A Wathel et al. 1998. Mol Cell, 1 : 507

Marshall C J. 1991. Tumor suppressor genes. Cell, 64: 313～326

Martorana A M, Zheng G, Springall F, Iland H I, O'Grady, Lyons J G. 1999. Absolute quantitation of Specific mRNA in cell and tissue samples by comparative PCR. Biotechmiques, 27: 136～144

Martorana H, Nirasawa S, Teraudi R. 1999. Transcript profiling in rice seedlings using sering analysis of gene expression(SAGE). The Plant journal, 20(6):719～726

Merrick W C, J W B Hershey. 1995. The pathway and mechanism of eukaryotic protein synthesis. In: Translational Control. USA: Cold Spring Harbor Laboratory Press

Mossi R, U Hübscher. 1998. Clamping down on clamps and clamp loaders-the eukaryotic repliction factor. C Eur J Biochem, 254: 209～216

Neer E J. 1995. Heterotrmeric G proteins: organizers of transmembrane signals. Cell, 80:249～257

Nicolas Pineaulta, Cheryl D, Helgason H Jeffrey. 2002. Differential expression of *Hox*, *Meis1*, and *Pbx1* genesin primitive cells throughout murine hematopoietic ontogeny. Lawrence Experimental Hematology, 30: 49～57

Pavlov M Y et al. 1997. Fast recycling of Escherichia coli ribosomes requires both ribosome recycling factor(RRF) and release factor RF3. EMBO J, 16: 4134～4141

Sachs A B, P Sarnow, M W Hentze. 1997. Starting at the beginning, middle, and end: translation initiation in eukaryotes. Cell, 89: 831～838

Sambrook J et al. 2001. Molecular cloning: A laboratory manual. 3rd ed. New York: Cold Spring Harbor labora-

tory Press. 1～560

Smit A F A. 1996. The origin of interspersed repeats in the human genome. Curr Opin Genet Dev, 6 : 743～748

Songyang Z et al. 1993. SH2 domains recognize specific phosphopeptide sequences. Cell, 74:1135～1145

Starlinger P. 1977. DNA rearrangments in prokaryotes. Ann Rev Genetics, 11: 103

UIlrich A, Schlessinger J. 1990. Signal transduction by receptors with tyrosine kinase activity. Cell, 61:203～212

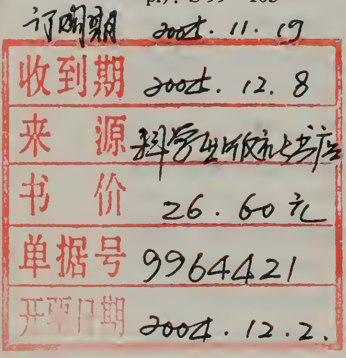
Wahle E, U Keuhn. 1997. The mechanism of 3′cleavage and polyadenylation of eukaryotic pre-mRNA . Prog Nucl Acid Res Mol Biol, 57: 41～71

West S C et al. 1981. Homologous pairing can occur before DNA strand separation in generalized genetic recombination. Nature, 290: 29

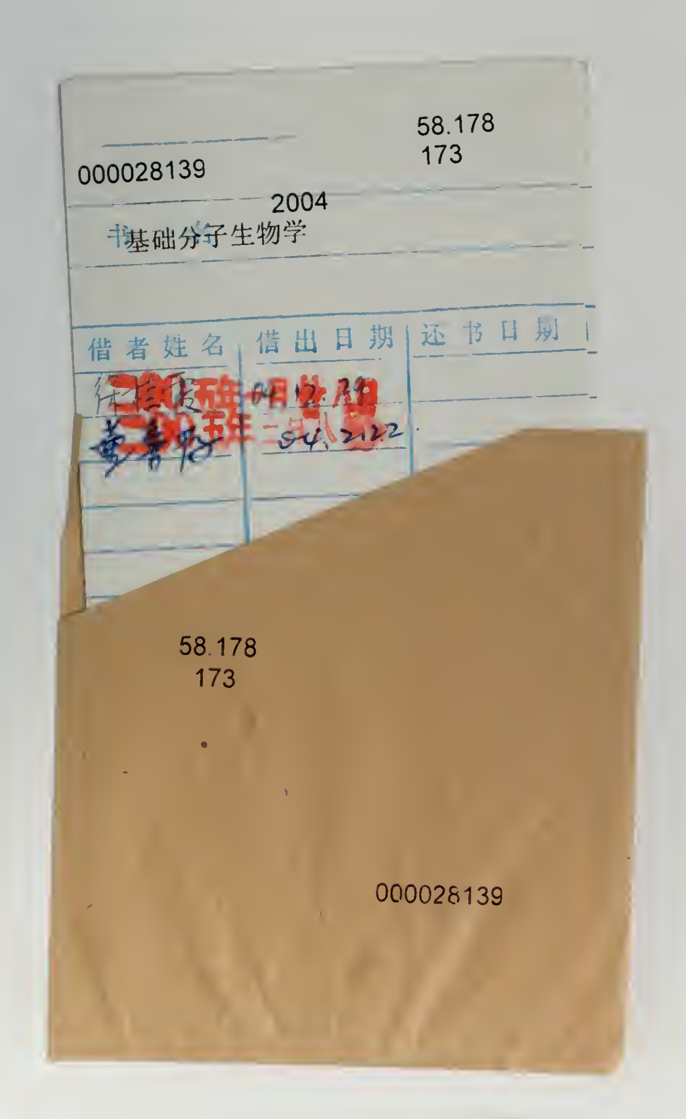
Wu R et al. 1989. Recombinant DNA Methodology. USA: Academic Press Inc

Yang Du, Janee L Campbell. 2002. Mapping gene expression patterns during myeloid differentiation using the EML hematopoietic progenitor cell line. Experimental Hematology, 30: 649～658

Zimmermann K C, Green D R. 2001. How cells die: apoptosis pathways. J Allergy Clin Immunol, 108 (4 suppl): S 99～103





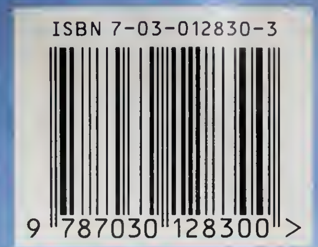


58.178
173
000028139
2004
基础分子生物学
借者姓名
借出日期
还书日期
58.178
173
000028139

(Q-1372.0101)

高等教育分社
联系电话：010-64019815
http://www.lifescience.com.cn
e-mail:bio-edu@cspg.net

基础分子生物学



ISBN 7-03-012830-3
定 价：38.00 元